

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2004-531207

(P2004-531207A)

(43) 公表日 平成16年10月14日(2004.10.14)

| | | |
|------------------------------------|----------------|-------------|
| (51) Int.Cl. ⁷ | F 1 | テーマコード (参考) |
| C 1 2 N 15/09 | C 1 2 N 15/00 | 4 B O 2 4 |
| A O 1 K 67/027 | A O 1 K 67/027 | 4 B O 2 9 |
| A 6 1 K 45/00 | A 6 1 K 45/00 | 4 B O 5 0 |
| A 6 1 P 43/00 | A 6 1 P 43/00 | 4 B O 6 3 |
| C O 7 K 16/40 | C O 7 K 16/40 | 4 B O 6 5 |
| 審査請求 未請求 予備審査請求 有 (全 193 頁) 最終頁に続く | | |

| | | | |
|---------------|------------------------------|----------|-----------------------|
| (21) 出願番号 | 特願2002-537892 (P2002-537892) | (71) 出願人 | 398021788 |
| (86) (22) 出願日 | 平成13年10月5日 (2001.10.5) | | ビーイー コーポレーション (エヌワイ) |
| (85) 翻訳文提出日 | 平成15年4月21日 (2003.4.21) | | アメリカ合衆国 カリフォルニア州 94 |
| (86) 国際出願番号 | PCT/US2001/042528 | | 404 フォスター シティ, リンカーン |
| (87) 国際公開番号 | W02002/034922 | | センター ドライブ 850 |
| (87) 国際公開日 | 平成14年5月2日 (2002.5.2) | (74) 代理人 | 100092901 |
| (31) 優先権主張番号 | 60/241,745 | | 弁理士 岩橋 祐司 |
| (32) 優先日 | 平成12年10月20日 (2000.10.20) | (72) 発明者 | メルクロフ, ゲナディ, ブイ |
| (33) 優先権主張国 | 米国 (US) | | アメリカ合衆国 メリーランド州 208 |
| (31) 優先権主張番号 | 09/739,456 | | 50 ロックビル, C2-4#21, ウェ |
| (32) 優先日 | 平成12年12月19日 (2000.12.19) | | スト グーデ ドライブ 45, セレーラ |
| (33) 優先権主張国 | 米国 (US) | | ジェニミクス内 |
| (31) 優先権主張番号 | 09/818,647 | | |
| (32) 優先日 | 平成13年3月28日 (2001.3.28) | | |
| (33) 優先権主張国 | 米国 (US) | | |
| 最終頁に続く | | | |

(54) 【発明の名称】 単離ヒト薬剤一代謝タンパク質、ヒト薬剤一代謝タンパク質をコード化する核酸分子及びその使用

(57) 【要約】

本発明は、ヒトゲノム中の遺伝子によりコード化されている、本発明の薬剤一代謝酵素ペプチドのアミノ酸配列を提供するものである。本発明は特に、単離ペプチド及び核酸分子、薬剤一代謝酵素ペプチドのオルトログ及びパラログの同定方法、及び薬剤一代謝酵素ペプチドのモジュレータの同定方法を提供するものである。

【特許請求の範囲】

【請求項1】

下記グループから選択されるアミノ酸配列から成る単離ペプチド。

- (a) SEQ ID NO. 2に示されるアミノ酸配列；
- (b) SEQ ID NO. 2に示されるアミノ酸配列の対立変異体のアミノ酸配列であって、該対立変異体は、SEQ ID NO. 1又は3に示される核酸分子の対向鎖に、ストリンジェントな条件下でハイブリダイズする核酸分子によってコード化されていることを特徴とするアミノ酸配列；
- (c) SEQ ID NO. 2で示されるアミノ酸配列のオルトログのアミノ酸配列であって、該オルトログは、SEQ ID NO. 1又は3に示される核酸分子の対向鎖に、ストリンジェントな条件下でハイブリダイズする核酸分子によってコード化されていることを特徴とするアミノ酸配列；及び
- (d) SEQ ID NO. 2に示されるアミノ酸配列のフラグメントであって、該フラグメントは、少なくとも10の隣接するアミノ酸を含むことを特徴とするアミノ酸配列。

10

【請求項2】

下記グループから選択されるアミノ酸配列を含む単離ペプチド。

- (a) SEQ ID NO. 2に示されるアミノ酸配列；
- (b) SEQ ID NO. 2に示されるアミノ酸配列の対立変異体のアミノ酸配列であって、該対立変異体は、SEQ ID NO. 1又は3に示される核酸分子の対向鎖に、ストリンジェントな条件下でハイブリダイズする核酸分子によってコード化されていることを特徴とするアミノ酸配列；
- (c) SEQ ID NO. 2に示されるアミノ酸配列のオルトログのアミノ酸配列であって、該オルトログは、SEQ ID NO. 1又は3に示される核酸分子の対向鎖に、ストリンジェントな条件下でハイブリダイズする核酸分子によってコード化されていることを特徴とするアミノ酸配列；及び
- (d) SEQ ID NO. 2に示されるアミノ酸配列のフラグメントであって、該フラグメントは、少なくとも10の隣接するアミノ酸を含むことを特徴とするアミノ酸配列。

20

【請求項3】

請求項2記載のペプチドに選択的に結合する単離抗体。

30

【請求項4】

下記グループから選択されるヌクレオチド配列から成る単離核酸分子。

- (a) SEQ ID NO. 2に示されるアミノ酸配列をコード化するヌクレオチド配列；
- (b) SEQ ID NO. 2に示されるアミノ酸配列の対立変異体のアミノ酸配列をコード化するヌクレオチド配列であって、該ヌクレオチド配列は、SEQ ID NO. 1又は3に示される核酸分子の対向鎖にストリンジェントな条件下でハイブリダイズすることを特徴とするヌクレオチド配列；
- (c) SEQ ID NO. 2に示されるアミノ酸配列のオルトログのアミノ酸配列をコード化するヌクレオチド配列であって、該ヌクレオチド配列は、SEQ ID NO. 1又は3に示される核酸分子の対向鎖にストリンジェントな条件下でハイブリダイズすることを特徴とするヌクレオチド配列；
- (d) SEQ ID NO. 2に示されるアミノ酸配列のフラグメントをコード化するヌクレオチド配列であって、該フラグメントは、少なくとも10の隣接するアミノ酸を含むことを特徴とするヌクレオチド配列；及び
- (e) (a)～(d)のヌクレオチド配列に相補的であるヌクレオチド配列。

40

【請求項5】

下記グループから選択されるヌクレオチド配列を含む単離核酸分子。

- (a) SEQ ID NO. 2に示されるアミノ酸配列をコード化するヌクレオチド配列；

50

(b) SEQ ID NO. 2 に示されるアミノ酸配列の対立変異体のアミノ酸配列をコード化するヌクレオチド配列であって、該ヌクレオチド配列は、SEQ ID NO. 1 又は 3 に示される核酸分子の対向鎖にストリンジェントな条件下でハイブリダイズすることを特徴とするヌクレオチド配列；

(c) SEQ ID NO. 2 に示されるアミノ酸配列のオルトログのアミノ酸配列をコード化するヌクレオチド配列であって、該ヌクレオチド配列は、SEQ ID NO. 1 又は 3 に示される核酸分子の対向鎖にストリンジェントな条件下でハイブリダイズすることを特徴とするヌクレオチド配列；

(d) SEQ ID NO. 2 に示されるアミノ酸配列のフラグメントをコード化するヌクレオチド配列であって、該フラグメントは、少なくとも 10 の隣接するアミノ酸を含むことを特徴とするヌクレオチド配列；及び

(e) (a) ~ (d) のヌクレオチド配列に相補的であるヌクレオチド配列。

【請求項 6】

請求項 5 記載の核酸分子を含む遺伝子チップ。

【請求項 7】

請求項 5 記載の核酸分子を含むヒト以外の遺伝子組み換え動物。

【請求項 8】

請求項 5 記載の核酸分子を含む核酸ベクター。

【請求項 9】

請求項 8 記載のベクターを含む宿主細胞。

【請求項 10】

請求項 1 記載の何れかのペプチドを製造する方法であって、(a) ~ (d) の何れかのアミノ酸配列をコード化するヌクレオチド配列を宿主細胞内に導入し、ペプチドがヌクレオチド配列から発現される条件下で宿主細胞を培養する方法。

【請求項 11】

請求項 2 記載の何れかのペプチドを製造する方法であって、(a) ~ (d) の何れかのアミノ酸配列をコード化するヌクレオチド配列を宿主細胞に導入し、ペプチドがヌクレオチド配列から発現される条件下で宿主細胞を培養する方法。

【請求項 12】

サンプル中における請求項 2 記載の何れかのペプチドの存在を検出する方法であって、サンプル中に該ペプチドの存在を特異的に検出する試薬とサンプルを接触させ、該ペプチドの存在を検出する方法。

【請求項 13】

サンプル中における請求項 5 記載の核酸分子の存在を検出する方法であって、ストリンジェントな条件下で該核酸分子にハイブリダイズするオリゴヌクレオチドとサンプルを接触させ、サンプル中の該核酸分子とオリゴヌクレオチドが結合するかどうかを判定する方法。

【請求項 14】

請求項 2 記載のペプチドのモジュレータを同定する方法であって、該ペプチドを試薬と接触させ、該試薬が該ペプチドの機能又は活性を変調したかどうかを判定する方法。

【請求項 15】

請求項 14 記載の方法において、前記試薬は前記ペプチドを発現する発現ベクターを含む宿主細胞に対して与えられる方法。

【請求項 16】

請求項 2 記載の何れかのペプチドに結合する試薬の同定方法であって、ペプチドと試薬を接触させ、接触混合物中にペプチドと試薬とが結合した複合体が形成されるかどうかをアッセイする方法。

【請求項 17】

請求項 16 記載の方法により同定された試薬と、薬学的に許容可能なそれらの担体とを含む薬剤組成物。

10

20

30

40

50

【請求項 18】

ヒト薬剤一代謝酵素タンパク質により媒介される疾患又は症状を治療する方法であって、請求項 16 記載の方法で同定された試薬を薬学的に有効な量、患者に投与する方法。

【請求項 19】

請求項 2 記載のペプチドの発現のモジュレータを同定する方法であって、該ペプチドを発現する細胞と試薬とを接触させ、該試薬が該ペプチドの発現を変調したかどうかを測定する方法。

【請求項 20】

SEQ ID NO. 2 に示されるアミノ酸配列と少なくとも 70% の相同性を持つアミノ酸配列を有する単離ヒト薬剤一代謝酵素ペプチド。

10

【請求項 21】

請求項 20 記載のペプチドにおいて、SEQ ID NO. 2 に示されるアミノ酸配列と少なくとも 90% の相同性を持つアミノ酸配列を有するペプチド。

【請求項 22】

ヒト薬剤一代謝酵素ペプチドをコード化している単離核酸分子であって、SEQ ID NO. 1 又は 3 に示される核酸分子と少なくとも 80% の相同性を有している核酸分子。

【請求項 23】

請求項 22 記載の核酸分子において、SEQ ID NO. 1 又は 3 に示される核酸分子と少なくとも 90% の相同性を有している核酸分子。

【発明の詳細な説明】

20

【0001】

技術分野

本発明は、ω-ヒドロキシラーゼ シトクロム P 450 薬剤一代謝酵素サブファミリーに関連した薬剤一代謝タンパク質、組み換え DNA 分子、及びタンパク質の製造に関する。本発明は、特に、ヒトの治療の開発に用いるための、新規な薬剤一代謝ペプチド、タンパク質、及びこれらのタンパク質分子をコード化する核酸分子を提供するものである。

【0002】

背景技術

薬剤代謝タンパク質

薬剤一代謝酵素（“DME”）の誘導は、生体異物に対する一般的な生物学的な応答であり、その機構及び結果は、学术界、産業界、及び薬理学、毒物学の規定領域において重要である。

30

【0003】

殆どの薬剤において、薬剤一代謝酵素が、どれだけの期間、どれだけの量、薬剤が体内に残存するのかを決定する。このため、薬剤の開発者達は、薬剤候補物とこれらの酵素との相互作用を特徴付けることの重要性を認識している。例えば、薬剤一代謝酵素 CYP2D6、シトクロム p 450（“CYP”）スーパーファミリーのメンバー、の多形性は、抗鬱薬、抗精神病薬、β遮断薬、及び抗不整脈薬を含む、広範囲にわたる薬剤の遅延又は超急速性の代謝物質産出する。このような薬剤代謝の異常割合は、薬剤の無効、又は全身への蓄積、及び毒性に至る可能性がある。

40

【0004】

候補薬剤を開発している薬理学者は、設計段階において、どの酵素が薬剤候補物を代謝するのか、及びその速度について、できるだけ早く知ることが重要である。歴史的に、薬剤の代謝経路上の酵素は、動物の代謝研究を通じて決定されてきたが、現在、このアプローチは主として、ヒトの組織の使用、又は薬剤代謝酵素のクローンの使用に取って代わり、これらの酵素の個別形態での特定の役割についての知見が与えられる。これらのツールを用い、薬剤候補物の定性的、及び定量的な成り行きを、最初にヒトに投与する前に予期することができる。結果として、代謝に要求される特性の選択及び最適化は、開発プロセスの初期に行うことが可能であり、このために、薬剤の臨床研究に続いて起こる予期できない毒性の問題、及び関連する諸費用を回避することができる。さらに、1つの薬剤

50

の他のものの性質への影響も推測することができる。

【0005】

既知の薬剤代謝酵素には、シトクロム p 4 5 0 (“ C Y P ”) スーパーファミリー、N-アセチルトランスフェラーゼ (“ N A T ”)、UDP-グルクロノシルトランスフェラーゼ (“ U G T ”)、メチルトランスフェラーゼ、アルコールデヒドロゲナーゼ (“ A L D H ”)、ジヒドロピリミジンデヒドロゲナーゼ (“ D P D ”)、N A D P H : キノンオキシドリダクターゼ (“ N O O ”、又は “ D T ジアホラーゼ ”)、カテコール O-メチルトランスフェラーゼ (“ C O M T ”)、グルタチオン S-トランスフェラーゼ (“ G S T ”)、ヒスタミンメチルトランスフェラーゼ (“ H M T ”)、スルホトランスフェラーゼ (“ S T ”)、チオプリンメチルトランスフェラーゼ (“ T P M T ”) 及びエポキシドヒドロキシラーゼが含まれる。薬剤代謝は、その代謝機能により、一般的に2つの相に分類される。相 I の酵素は官能基の修飾に触媒作用を及ぼし、相 I I の酵素は内生の置換基との結合に触媒作用を及ぼす。これらの分類は、他の薬剤代謝の機構が発見されているように、排他的又は完全なものと解釈するべきではない。例えば、活性な輸送機構の使用が、解毒のプロセスの一部として特徴づけられている。

10

【0006】

相 I の反応には、アミナーゼの脱アミノ化、エステル及びアミドの加水分解、例えば、グリシン及びリン酸塩との結合反応、シトクロム p 4 5 0 酸化／還元酵素システムによる酸化、及び脂肪酸経路における分解のような、分解作用プロセスが含まれる。加水分解は、種々の非-特異性ハイドrolラーゼ及びエステラーゼにより、主に肝臓、及び血漿内で起こる。デアミナーゼとアミダーゼの両者は、共に肝臓及び血清中に局所化し、分解作用プロセスの大部分を行っている。還元反応は、主に小胞体の細胞内で起こる。

20

【0007】

相 I I の酵素は、毒性物質と水溶性物質との結合に触媒作用を及ぼし、これによって毒性物質の水溶性が増し、排出速度が増大することによって、毒性物質を無毒化する。さらに、結合は毒の生化学的反応性を減少させる。相 I I の酵素の例には、それぞれ、グルタチオン、グルクロン酸の結合を触媒する、グルタチオン-S-トランスフェラーゼ、及びUDP-グルクロノシルトランスフェラーゼが含まれる。トランスフェラーゼは、主に腎臓及び肝臓において結合反応を行う。

【0008】

肝臓は、精神作用性薬剤を含む殆どの薬剤の除去を行う主要な部位であって、それぞれ薬剤を酸化、又は結合している相 I 及び相 I I の酵素の両者が含まれる。

30

【0009】

医師達は、現在、薬剤を人口の平均に基づいた投与量処方しており、遺伝的な変異性は考慮に入れられていない。薬剤代謝における個人間の変異性は、通常、遺伝的及び環境的因子の両者、特に薬剤-代謝酵素がどのように制御されているかによるものである。ある酵素では、遺伝子成分が優勢であり、変異性は、標準の、野生一型の酵素の変異体に関係している。

【0010】

多くの薬剤-代謝酵素は、臨床に関連した遺伝子多形性を示す。本質的に、官能基の修飾又は内生置換基の結合を担っている、全ての主なヒト酵素は、一般的に遺伝子レベルでの多形を示す。例えば、非-機能性変異体酵素を発現する多形は、より薬剤の濃度に依存した影響を受けやすい傾向にある患者のサブグループを生じる。最近の遺伝子型分類の開発は、影響を受ける個人の特定を可能にする。この結果、異常な代謝、影響を受けた酵素による薬剤代謝への応答が理解及び予期され、このため、医師達が、改良された治療法を実行するために、薬剤の投与量を調整することが可能となる。

40

【0011】

同様のアプローチもまた、種々の癌の発達に関連した危険因子の同定において重要になっている。これは、薬剤代謝に関連した酵素が、化学発癌性物質の活性化、解毒作用をも担っているためである。具体的には、腫瘍形成の発達は、発癌性物質を活性化する相 I 酵素

50

と、それらが無毒化する相ⅠⅠ酵素とのバランスによって調整される。したがって、したがって、個人の癌感染性はしばしばこれらの2つのプロセスの間のバランスに関連しており、それは部分的に、遺伝的に決定されており、適当な遺伝子型テストによってスクリーンされることができる。相ⅠⅠ酵素と比較した相Ⅰ酵素のより高い誘導は、大量の求電子試薬及び反応性酸素種を生じ、DNA及び膜の損傷、及び他の腫瘍形成にいたる副作用を引き起こすかもしれない。逆に、相ⅠⅠ酵素のより高いレベルでの発現は、細胞を種々の化合物から保護することを可能とする。

【0012】

薬剤一代謝酵素の異常な活性は、癌、パーキンソン病、筋緊張性ジストロフィー、及び発展性の障害を含む、ヒトの疾患に関連している。

10

【0013】

シトクロム p 4 5 0

相Ⅰの薬剤一代謝酵素の例として、シトクロム p 4 5 0 (“CYP”) スーパーファミリーのメンバーが挙げられ、このメンバーには、肝臓において発現される主要な薬剤代謝酵素が含まれる。CYPスーパーファミリーは、多くのタイプの化学反応を触媒的作用を及ぼす、何百といった多くの種のイソ型を持ち、その機能において広大な多様性を持つ。CYPスーパーファミリーには、少なくとも30の関連した酵素が含まれ、これらはアミノ酸相同性によって異なるファミリーに分類される。CYPファミリーの例としては、薬剤及び生体異物の代謝を担っている小胞体タンパク質を含んだCYPファミリー1, 2, 3, 4が含まれる。これら4つのファミリーのうちの、約10~15の個々の遺伝子生成物は、何千もの多様な構造の化合物を代謝する。CYPスーパーファミリーの酵素は、総合して、ヒトに用いられ入手可能な全ての薬剤のうちの80%以上の代謝に関係していると思われられる。例えば、CYP1Aサブファミリーには、アセトアミノフェン、アミトリプチリン、カフェイン、クロザピン、ハロペリドール、イミプラミン、オランザピン、オンダンセトロン、フェナセチン、プロパフェノン、プロプラノロール、タクリン、テオフィリン、ベラパミルを含む、広範囲にわたって使用される幾つかの薬剤を代謝するCYP1A2が含まれる。さらに、CYP酵素は、プロスタグランジン及びステロイドを含むいくつかの内生酵素の代謝において、付加的な役割を果たす。

20

【0014】

いくつかのCYP酵素が多形性の形態で存在しており、これは人口のうちのわずかな割合が、通常は、活性の減少又は停止によって、酵素の活性が変化する変異遺伝子を有していることを意味している。例えば、遺伝子の多形性は、CYP2C19及びCYP2D6遺伝子において良く特徴づけられている。CYP2C19の基質には、クロミプラミン、ジアゼパム、イミプラミン、メフェニトイン、モクロベミド、オメプラゾール、フェニトイン、プロプラノール、及びプロパフェノンが含まれる。これらの遺伝子の多形変異体は、これらの基質を異なる割合で代謝し、これは患者への有効な治療の投薬において効果的な作用を及ぼすことができる。

30

【0015】

CYPの基質特異性は、これらの合成物の全ての代謝に適応するために非常に広くなければならないとはいえ、個々のCYP遺伝子生成物のそれぞれは、その結合及び触媒部位によって規定されるよりもより狭い基質特異性を有している。薬剤代謝は、特定のCYP遺伝子生成物の量又は活性の変化により調節されることができる。CYP調節の方法としては、CYPの発現における遺伝子変化（例えば、遺伝子多形性）、CYPと結合可能な他の生体異物によるCYP代謝の阻害、及び薬剤自身又は他の生体物による特定のCYPの誘導が含まれる。CYPの抑制及び誘導は、逆向きの薬剤相互作用の最も一般的な機構である。例えば、CYP3Aサブファミリーは、心律動異常を生じる非沈静化坑ヒスタミン剤及びシサプリドに関連した、臨床的に重要な薬剤相互作用に関連している。他の例では、CYP3A4及びCYP1A2酵素は、多くの精神療法剤の代謝を担っている。さらに、CYP酵素は、ヒト免疫不全ウイルスに感染した患者の治療に用いられるプロテアーゼ抑制剤を代謝する。これらの酵素の特異な機能と特徴を理解することによって、医師達は

40

50

薬剤の相互作用を予期、及び処理することができるかもしれず、また、特定の治療法に対する個人の応答を予測、又は説明することができるかもしれない。

【0016】

CYPスーパーファミリーに触媒される反応の例としては、過酸化物を、ヒドロキシル化反応における酸素ドナー、還元 β 分裂の基質、アルデヒドからギ酸塩及びアルケンへの開裂におけるペルオキシヘミアセタール中間体として用いた過酸化反応が含まれる。脂質ヒドロペルオキシドは、炭化水素及びアルデヒド酸を得るために、還元 β 開裂を受ける。これらの生成物の1つ、トランス-4-ヒドロキシノネナールは、消極的な調節プロセスであるかもしれないCYP、特にアルコール誘導性2E1を不活性化する。CYP鉄-オキセン種は、殆どのヒドロキシル化反応の酸素ドナーであると考えられているが、鉄-ペルオキシ種は、芳香族化反応のように、残存する構造の不飽和化を伴う多くのアルデヒドの脱ホルミル化に関連していると考えられる。

10

【0017】

CYPに関連した酸化型代謝の薬剤の例としては、アセトアミノフェン、アルフェンタニル、アルプラゾラム、アルプレノロール、アミオダロン、アミトリプチリン、アステミゾール、ブスピロン、カフェイン、カルバマゼピン、クロルフェニラミン、シサプリド、クロミプラミン、クロミプラミン、クロザピン、コデイン、コルヒチン、コルチゾール、シクロホスファミド、サイクロスポリン、ダブソン、デシプラミン、デキストロメトर्फアン、ジアゼパム、ジクロフェナク、ジルチアゼム、エンカイニド、エリスロマイシン、エストラジオール、フェロジピン、フルオキセチン、フルバスタチン、ハロペリドール、イブプロフェン、イミプラミン、インジナビル、インドメタシン、インドラミン、イルベサルタン、リドカイン、ロサルタン、マクロライド抗生物質、メフェニトイン、メサドン、メトプロロール、メキシレチン、ミダゾラム、モクロベミド、ナプロキセン、ネファゾドン、ニカルジピン、ニフェジピン、ニトレンジピン、ノルトリプチリン、オランザピン、オメプラゾール、オndanセトロン、オキシコドン、パクリタキセル、パロキセチン、フェナセチン、フェニトイン、ピロキシカム、プロゲステロン、プロパフェノン、プロプラノロール、キニジン、リトナビル、サキナビル、セルトラリン、シルデナフィル、S-ワルファリン、タクリン、タモキシフェン、テノキシカム、テルフェナジン、テストステロン、テオフィリン、チモロール、トルブタミド、トリアゾラム、ベラパミル、ビンブラスチンが含まれる。

20

30

【0018】

相Iの酵素の異常活性は、ヒトの疾患の範囲に関連している。例えば、増大されたCYP2D6の活性は、膀胱、肝臓、咽頭、胃、及び肺の悪性腫瘍と関連しているのに対して、減少されたCYP2D6の活性は、パーキンソン病の危険性の増大につながる。CYPスーパーファミリーの欠損と関連した発達上の障害には、脳髄黄色腫症、副腎過形成、女性化乳房症、及び筋緊張性ジストロフィーが含まれる。

【0019】

ω -ヒドロキシラーゼ シトクロムP450

本発明により提供される新規なヒトタンパク質、およびそれをコード化する遺伝子は、例えば、シトクロムP4504A4 (CYP4A4)、シトクロムP-450p-2、プロスタグランジン、 ω -ヒドロキシラーゼ、及びラウリン酸 ω -ヒドロキシラーゼを含む、 ω -ヒドロキシラーゼシトクロムP450ファミリーに関連している。 ω -ヒドロキシラーゼシトクロムP450タンパク質は、プロスタグランジンA、及び、カプリン酸塩、ラウリン酸塩、パルミチン酸塩のような脂肪酸の ω - (ω 1を含む) ヒドロキシル化に触媒作用を及ぼす (Yoshimura et al., J Biochem (Tokyo) 1990 Oct; 108 (4): 544-8)。CYP4A4は、妊娠期間中に増大させられる (Palmer et al., Arch Biochem Biophys 1993 Feb 1; 300 (2): 670-6)。

40

【0020】

Matsubara et al., J Biol Chem 1987 Sep 25; 26

50

2 (27) : 13366-71; Yamamoto et al., (1984) J. Biochem. (Tokyo) 96, 593-603; Yokotani et al., Eur J Biochem 1991 Mar 28; 196 (3) : 531-6; and Johnson et al., Biochemistry 1990 Jan 30; 29 (4) : 873-9.

【0021】

本発明により提供されるタンパク質のようなシトクロムは、上述した点に加えて、多くの有用性を持っている。シトクロムは通常の生理学的基質を代謝するだけでなく、環境中の毒を中和する。肝細胞中のステロイド、脂肪酸、異物の酸化に加えて、シトクロムは、毒性化合物、殺虫剤、発癌性物質により誘導されることもできる。

10

【0022】

シトクロムの免疫学的及びPCRベースのアッセイは、実験医学における毒性、及び回転率の判定に用いられ得る。選択的な細胞毒薬剤は、特定のシトクロムと相互作用し、細胞死を誘引するように設計することができ、これによって潜在的な癌の新しい治療方法が提供される。

【0023】

シトクロムは、心筋細胞及び血管内皮細胞の損傷を引き起こすフリーラジカルを生成することができる。実験モデルでは、 α -トコフェロール、及び他の酸化防止剤は、フリーラジカルの生成を抑制する。グルタチオン及びグルタチオンペルオキシダーゼは、フリーラジカル-誘発性の細胞損傷からの自然保護に寄与している。全てのシトクロムを特徴づけることは、より効果的な酸化防止剤の開発を助長することとなる。本発明により提供される配列は、特定の化学予防薬剤の設計に用いることができる。

20

【0024】

ここに提供されるシトクロムは、他のヒトシトクロムと同様に、非-ヒトオキシゲナーゼを阻害し、ヒト酵素に対して毒性を示さない非-寄生性薬剤の発見のためのハイスループット薬剤スクリーンに用いることができる。

【0025】

CYPスーパーファミリーの更なる説明は、Igarashi et al., Arch Biochem Biophys 1997 Mar 1; 339 (1) : 85-91; Med Lett Drugs Ther 2000 Apr 17; 42 (1076) : 35-6 (no authors listed); Fowler et al., Biochemistry 2000 Apr 18; 39 (15) : 4406-14; Lamb et al., Chem Biol Interact 2000 Mar 15; 125 (3) : 165-75; Chiba et al., Xenobiotica 2000 Feb; 30 (2) : 117-29; and Meehan et al., Am J Hum Genet 1988 Jan; 42 (1) : 26-37を参照されたい。

30

【0026】

CYPスーパーファミリーは、薬剤の作用及び開発の主要なターゲットである。したがって、従来未知のCYPスーパーファミリーのメンバーを同定し特徴づけることは、薬剤開発の分野において有用である。

40

【0027】

UDP-グルクロノシルトランスフェラーゼ

相I代謝に関連した潜在的な薬剤相互作用への認識は高まっている。薬剤代謝に関連する相I酵素の重要なグループは、グルクロノシルトランスフェラーゼ、特にUDP-グルクロノシルトランスフェラーゼ ("UGT") スーパーファミリーである。UGTスーパーファミリーのメンバーは、脂溶性化合物への糖ドナーとしてのUDPグルクロン酸の酵素添加に触媒作用を及ぼし、これらの水への可溶性を増大し、排出率を増大するプロセスである。哺乳類において、グルクロン酸は、代謝の廃棄物、及び環境から体内に入って毒性レベルとなる脂溶性化合物の蓄積の予防に用いられている。グルクロノシルトランスフェラーゼの誘導剤及び阻害剤の両者は公知であり、血漿濃度、及び向精神薬を含む重要

50

な薬剤の作用に影響する可能性を持っている。

【0028】

UGTスーパーファミリーは、いくつかの種の酵素のいくつかのファミリーを含んでおり、CYPスーパーファミリーのメンバーを定義するのに用いられるのに類似した命名法で定義される。動物、酵母、植物、及び細菌においては、少なくとも110の異なるCYPスーパーファミリーのメンバーが知られている。ヒトにおいては、33程度のファミリーが定義され、3つのファミリーが同定された。異なるUGTファミリーでは、アミノ酸配列の相同性が45%未満であるとして定義され；サブファミリーの中では約60%の相同性がある。UGTスーパーファミリーのメンバーはさらに、動物、植物、細菌において見出されるUDPグリコシルトランスフェラーゼスーパーファミリーの一部である。

10

【0029】

相II酵素、特にUGT酵素の役割は、精神薬理学において重要なものとして、認識が高まっている。UGT酵素は、多くの重要な向精神薬を変化させ、薬剤応答及び薬剤相互作用の変化の主要な原因である。例えば、ベンゾジアゼピン、ロラゼパム、オキサゼパム、及びテマゼパムは、尿中に排出される前にのみ、相II反応を受ける。

【0030】

相II酵素は、発癌性物質のような危険性物質を代謝し、解毒する。相II酵素をコード化する遺伝子の発現は、多くの薬剤により上方調整されることが知られている。例えば、オルチプラズは、相II酵素発現を上方調整する。研究により、選択性相II酵素誘導剤を発癌物質の前に投与した場合の、発癌性物質の癌誘因効果からの保護が証明された。発癌性物質の暴露に関連したヒトの癌の予防における、相II酵素誘導剤の潜在的な用途は、これらの分子の影響の理解を目標とした研究を促すこととなった。現在の生化学及び分子生物学の研究の方法を、選択性相II酵素誘導剤及びそのターゲットを同定、及び特徴づけるために用いることができる。癌の化学予防薬剤に応答する遺伝子の同定は、それらの基本的なメカニズムの研究を容易にし、遺伝子調節、酵素多形性、及び発癌性物質の解毒の関係についての知見を提供するものである。

20

【0031】

UGT酵素に関連した代謝を変化させる薬剤の例としては、アミトリプチリン、ブプレノルフィン、クロルプロマジン、クロザピン、コデイン、シプロヘプタジン、ジヒドロコデイン、ドキシセピン、イミプラミン、ラモトリジン、ロラゼパム、モルヒネ、ナロルフィン、ナルトレキソン、テマゼパム、及びバルプロエートが含まれる。

30

【0032】

相IIの酵素の異常活性は、ヒトの疾患の範囲に関連している。例えば、ギルバート症候群は、UGT1遺伝子の変異により起こる常染色体優勢障害であり、UGT1A1酵素の変異は、クリグラナーナジャー症候群に関係していることが証明された。

【0033】

UGTスーパーファミリーは、薬剤の作用及び開発の主要なターゲットである。したがって、従来未知のUGTスーパーファミリーのメンバーを同定し特徴づけることは、薬剤開発の分野において有用である。

【0034】

薬剤一代謝酵素、特にω-ヒドロキシラーゼシトクロムP450薬剤一代謝酵素サブファミリーのメンバーは、薬剤の作用及び開発の主要なターゲットである。したがって、この従来未知の薬剤一代謝タンパク質のサブファミリーのメンバーを同定し特徴づけることは、薬剤開発の分野において有用である。本発明は、ω-ヒドロキシラーゼシトクロムP450薬剤一代謝酵素サブファミリーのメンバーとの相同性を持った、従来未確認のヒト薬剤一代謝タンパク質を提供することによって、これらの技術の進展に寄与するものである。

40

【0035】

発明の要約

本発明は、ω-ヒドロキシラーゼシトクロムP450薬剤一代謝酵素サブファミリー、

50

これらの対立遺伝子変異体、および他の哺乳類におけるこれらのオルトログに関係する、ヒト薬剤一代謝酵素ペプチド及びタンパク質のアミノ酸配列同定の一部に基づいたものである。これらの特異なペプチド配列、及びこれらのペプチドをコード化する核酸配列は、ヒトの治療ターゲット開発のためのモデルとして用いることができ、治療に用いるタンパク質の識別に役立ち、また、薬剤一代謝酵素を発現する細胞及び組織内での薬剤一代謝酵素活性を調節するようなヒトの治療薬の開発におけるターゲットとなり得る。図1の実験データによると、ヒトの胃、脳（乳児を含む）、子宮内膜腫瘍、前立腺、腎臓、副腎腫瘍、頭部／頸部、交感神経幹、胸部、及び肝細胞腫瘍で発現することが示されている。

【0036】

発明の詳細な説明

10

概論

本発明は、ヒトゲノムの配列解析に基づいている。ヒトゲノムの配列解析及び構築に際して、配列情報を解析することによって、当該分野において、薬剤一代謝酵素タンパク質又はその一部であると同定及び特徴付けられ、ω-ヒドロキシラーゼ シトクロム P 4 5 0 薬剤一代謝酵素サブファミリーに関連付けられるタンパク質／ペプチド／ドメインに対して、構造及び／又は配列の相同性を有するペプチドをコード化している、ヒトゲノムの従来未確認のフラグメントが明らかとなった。これらの配列を用いて、付加的なゲノム配列を構築し、転写及び／又は cDNA 配列を単離し、特徴付けた。この解析に基づいて、本発明は、ω-ヒドロキシラーゼ シトクロム P 4 5 0 薬剤一代謝酵素サブファミリーに関連するヒト薬剤一代謝酵素ペプチド及びタンパク質のアミノ酸配列、これらの薬剤一代謝酵素ペプチド及びタンパク質をコード化する転写形態の核酸配列、cDNA 配列及び／又はゲノム配列、核酸変異（対立遺伝子情報）、発現の組織分布、及び本発明の薬剤一代謝酵素に対して構造又は配列の相同性を有する、最も関連性の高い既知のタンパク質／ペプチド／ドメインに関する情報を提供するものである。

20

【0037】

本発明において提供されるペプチドは、従来未知であることに加えて、商業的に重要な製品及びサービスの開発において有用であるという点に基づいて選択され得るものである。特に、本発明のペプチドは、ω-ヒドロキシラーゼ シトクロム P 4 5 0 薬剤一代謝酵素サブファミリーにおける既知の薬剤一代謝酵素タンパク質、及びその発現パターンに対して相同性及び／又は構造上の相関性を有していることに基づいて選択される。図1の実験データによると、ヒトの胃、脳（乳児を含む）、子宮内膜腫瘍、前立腺、腎臓、副腎腫瘍、頭部／頸部、交感神経幹、胸部、及び肝細胞腫瘍で発現することが示されている。この技術は、本発明の遺伝子と類似した発現パターンを有するこのファミリーのタンパク質及びペプチドにおける商業的重要性を確立するものである。本発明のペプチドのより具体的な性質及びその使用に関しては、本明細書、特に背景技術、図面の注釈に記載され、及び／又は薬剤一代謝酵素タンパク質の既知のω-ヒドロキシラーゼ シトクロム P 4 5 0 薬剤一代謝酵素サブファミリー又はサブファミリーのそれぞれにおいて周知である。

30

【0038】

【実施例の詳細】

ペプチド分子

40

本発明は、薬剤一代謝酵素ファミリーのタンパク質のメンバーであると同定されるタンパク質分子をコード化する核酸配列を提供するものであり、これらはω-ヒドロキシラーゼ シトクロム P 4 5 0 薬剤一代謝酵素サブファミリーに関連付けられる（図2にタンパク質配列、図1に転写／cDNA配列、図3にゲノム配列を示す）。図2に示されるペプチド配列、同様に本明細書に記載される対立変異体、特に本明細書及び図3の情報を用いて同定される対立変異体のような明らかな変異体は、ここでは本発明の薬剤一代謝酵素ペプチド、薬剤一代謝酵素ペプチド、又は本発明のペプチド／タンパク質と呼ばれる。

【0039】

本発明は、図2に示される薬剤一代謝酵素ペプチドのアミノ酸配列から成る、あるいは実質的に成る、あるいはこれを含む単離ペプチド及びタンパク質分子を提供する（図1に示

50

される核酸分子である転写／cDNA、又は図3に示されるゲノム配列によりコード化される)とともに、本技術により調製、使用されるこれらのペプチドの明らかな変異体を提供するものである。これらの変異体については以下で詳述する。

【0040】

ここで用いられる場合、ペプチドが細胞物質を実質的に含まない、又は化学前駆物質あるいは他の化学物質を含まない場合に、ペプチドは「単離」又は「精製」されたという。本発明のペプチドは、同質又は他の純度になるまで精製することができる。精製のレベルは使用目的に基づく。重要な性質としては、調製物中に他の成分が多量に存在していたとしても、要求されるペプチドの機能を発揮できるということである(単離核酸分子の性質については後述する)。

10

【0041】

いくつかの使用では、「実質的に細胞物質を含まない」とは、他のタンパク質(すなわち汚染タンパク質)を約30%(乾燥重量)未満、他のタンパク質を約20%未満、他のタンパク質を約10%未満、又は、他のタンパク質を約5%未満有するペプチド調製物を含む。ペプチドが組み換えにより製造される場合、培地がタンパク質調製物の容量に対して20%未満のときには、実質的に培地は存在しないとしてもできる。

【0042】

「実質的に化学前駆物質又は他の化学物質を含まない」という用語は、合成に関与した化学前駆物質又は他の化学物質から分離されたペプチド調製物をいう。ある例においては、「実質的に化学前駆物質又は他の化学物質を含まない」とは、化学前駆物質又は他の化学物質を約30%(乾燥重量)未満、化学前駆物質又は他の化学物質を約20%未満、化学前駆物質又は他の化学物質を約10%未満、又は、化学前駆物質又は他の化学物質を約5%未満有する薬剤一代謝酵素ペプチド調製物を含む。

20

【0043】

単離薬剤一代謝酵素ペプチドは、自然に発現する細胞、発現のために変形された(組み換え)細胞から精製するか、又は、既知のタンパク質合成方法を用いて合成することができる。図1に示す実験データは、図1の実験データによると、ヒトの胃、脳(乳児を含む)、子宮内膜腫瘍、前立腺、腎臓、副腎腫瘍、頭部/頸部、交感神経幹、胸部、及び肝細胞腫瘍で発現することが示されている。例えば、薬剤一代謝酵素ペプチドをコード化している核酸分子は、発現ベクター中にクローニングされ、さらにこの発現ベクターは宿主細胞に導入されて、タンパク質が宿主細胞内で発現する。そして、タンパク質は標準のタンパク質精製技術を用いた適当な精製スキームによって、細胞から単離することができる。これら多くの技術については、以下で詳述する。

30

【0044】

したがって、本発明は、図2に示されるアミノ酸配列(SEQ ID NO. 2)から成るタンパク質、例えば、図1に示される転写／cDNA核酸配列(SEQ ID NO. 1)、及び図3に示されるゲノム配列(SEQ ID NO. 3)によりコード化されるタンパク質を提供するものである。このようなタンパク質のアミノ酸配列を図2に示す。このアミノ酸配列がタンパク質の最終的なアミノ酸配列であるとき、タンパク質はアミノ酸配列から成る。

40

【0045】

本発明はさらに、図2に示されるアミノ酸配列(SEQ ID NO. 2)から実質的に成るタンパク質、例えば、図1に示される転写／cDNA核酸配列(SEQ ID NO. 1)、及び図3に示されるゲノム配列(SEQ ID NO. 3)によりコード化されるタンパク質を提供するものである。このようなアミノ酸配列に微量の付加アミノ酸残基、例えば、最終的にタンパク質中に約1~100程度の付加残基、典型的には1~20の付加残基が存在する場合、タンパク質はアミノ酸配列から実質的に成る。

【0046】

本発明はさらに、図2に示されるアミノ酸配列(SEQ ID NO. 2)を含むタンパク質、例えば図1に示される転写／cDNA核酸配列(SEQ ID NO. 1)、及び図3

50

に示されるゲノム配列 (SEQ ID NO. 3) によりコード化されるタンパク質を提供するものである。このアミノ酸配列が、タンパク質の最終的なアミノ酸配列の少なくとも一部である場合、タンパク質はアミノ酸配列を含む。このような場合、タンパク質はペプチドのみであるか、又はタンパク質と自然に結びついているアミノ酸残基、あるいはアミノ酸残基／ペプチド配列に非相同のアミノ酸残基のような、付加的なアミノ酸分子 (コード化された配列と隣接する) を有することができる。このようなタンパク質は、少量の付加的アミノ酸残基を有するか、又は数百あるいはそれ以上の付加的アミノ酸を含むことができる。本発明の薬剤一代謝酵素ペプチドが含まれるタンパク質の好適な種として、天然の成熟タンパク質がある。これらの各種タンパク質を調製／単離する方法については、以下に簡単に述べる。

10

【0047】

本発明の薬剤一代謝酵素ペプチドは、キメラ又は融合タンパク質を形成するために、非相同性の配列に結合することができる。このようなキメラ及び融合タンパク質は、薬剤一代謝酵素ペプチドに対して実質的に相同性のないアミノ酸配列を有する非相同タンパク質に、有効に結合される薬剤一代謝酵素ペプチドを含む。「有効に結合される」とは、薬剤一代謝酵素ペプチドと非相同タンパク質がフレーム中で融合していることを示す。非相同タンパク質は、薬剤一代謝酵素ペプチドのN末端又はC末端に融合されることができる。

【0048】

いくつかの使用では、融合タンパク質は、薬剤一代謝酵素ペプチド自体の活性に影響を及ぼさない。融合タンパク質には、例えば、 β ガラクトシダーゼ融合、イースト2-ハイブリッドGAL融合、ポリHis融合、MYC標識、HI標識及びIg融合といった酵素融合タンパク質が含まれるが、これに限られるものではない。このような融合タンパク質、特にポリHis融合は、組み換え薬剤一代謝酵素ペプチドの精製に有用である。ある種の宿主細胞 (例えば哺乳類の宿主細胞) においては、タンパク質の発現及び／又は分泌は、非相同信号配列を用いることにより増加させることができる。

20

【0049】

キメラ又は融合タンパク質は、標準の組み換えDNA技術により製造することができる。例えば、異なるタンパク質配列をコードしているDNAフラグメントは、従来技術に従ってフレーム中に共に配置される。他の例では、融合遺伝子は、自動DNA合成機を含む従来技術により合成することが可能である。あるいは、遺伝子フラグメントのPCR増幅にアンカープライマーを用い、2つの連続的な遺伝子フラグメント間に相補的な突出部を形成し、その後、アニールすることにより、キメラ遺伝子配列を再増幅することができる (Ausubel et al., Current Protocols in Molecular Biology, 1992参照)。さらに、発現ベクターとしては、既に融合部分 (例えばGSTタンパク質) をコード化した多くのものを、商業的に入手することができる。薬剤一代謝酵素ペプチドをコード化した核酸は、融合部がフレーム中で薬剤一代謝酵素ペプチドに結合するようにして、このような発現ベクター中にクローニングされることができる。

30

【0050】

以上説明したように、本発明はまた、天然のペプチド成熟形態、ペプチドの対立遺伝子／配列変異体、非天然の組換え誘導変異体、ペプチドのオルトログ及びパラログなど、本発明のタンパク質のアミノ酸配列における明らかな変異体を提供、及び実施可能にするものである。このような変異体は、核酸組み換え技術やタンパク質生化学の分野で公知の技術を用いることにより、容易に生成することができる。しかしながら、この変異体には、本発明以前に公開されている何れのアミノ酸配列も含まれないものであると理解される。

40

【0051】

このような変異体は、ここに示される分子技術や配列情報を用いることにより、容易に同定／製造することが可能である。さらに、このような変異体は、本発明の薬剤一代謝酵素ペプチドに対する配列及び／又は構造上の相同性に基づいて、他のペプチドと容易に区別することができる。相同性／同一性の程度は、主に、ペプチドが機能的変異体であるか又

50

は非機能的変異体であるか、パラログファミリー中に存在する分化量、オルトログ間の進化的相違に基づいている。

【0052】

2つのアミノ酸配列、又は2つの核酸配列の同一性パーセントを決定するために、最適な比較を行う目的で、配列はアラインされる（例えば、最適なアラインメントのために、ギャップが第一及び第二アミノ酸又は核酸配列の一方又は両方に導入されることができ、非相同性配列は比較を行うために無視することができる）。好適な例としては、対象配列の長さの少なくとも30%、40%、50%、60%、70%、80%、90%又はそれ以上が、比較目的に応じてアラインされる。そして、対応するアミノ酸配置又はヌクレオチド配置上の、アミノ酸残基又はヌクレオチドが比較される。第一配列での配置が、第二配列において対応する配置と同じアミノ酸残基又はヌクレオチドによって占められているとき、分子はその配置上で同一である（ここで用いられているアミノ酸又は核酸の「同一性」は、アミノ酸又は核酸の「相同性」と同意義である）。2つの配列間の同一性パーセントは、配列において共有される同一配置数の関数であり、ギャップ数及び各ギャップ長さを考慮し、2つの配列の最適なアラインメントのために導入される必要がある。

【0053】

配列の比較、及び2つの配列間における同一性、類似性パーセントの決定は、数学的アルゴリズムを用いて行うことができる（Computational Molecular Biology, Lesk, A. M., ed., Oxford University Press, New York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D. W., ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part 1, Griffin, A. M., and Griffin, H. G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; and Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991）。好適な例としては、2つのアミノ酸配列間の同一性パーセントは、GCGソフトウェアパッケージ（<http://www.gcg.com>で入手可能）のGAPプログラムに組み込まれたNeedleman・Wunschアルゴリズム（J. Mol. Biol. (48): 444-453 (1970)）を用い、Blossom62マトリックス、又はPAM250マトリックス、及びギャップ重量16、14、12、10、8、6、4、長さ重量1、2、3、4、5、6を用いて決定される。他の好適な例としては、2つのヌクレオチド配列間の同一性パーセントは、GCGソフトウェアパッケージ（<http://www.gcg.com>で入手可能）のGAPプログラム（Devereux, J., et al., Nucleic Acids Res. 12(1): 387 (1984)）を用い、NW S gap dna. CMPマトリックス、ギャップ重量40、50、60、70、80、長さ重量1、2、3、4、5、6を用いて決定される。さらに他の一例としては、2つのアミノ酸又はヌクレオチド配列間の同一性パーセントは、ALIGNプログラム（バージョン2.0）に組み込まれたE. Myers・W. Millerのアルゴリズム（CABIOS, 4: 11-17 (1989)）を用い、PAM120重量残基表、ギャップ長ペナルティ12、ギャップペナルティ4を用いて決定される。

【0054】

本発明の核酸及びタンパク質配列は、例えば他のファミリー又は関連した配列を発見するために、配列データベースに対して検索を行う「クエリー配列」として用いられることができる。このような検索は、AltschulらのNBLAST、及びXBLASTプログラム（バージョン2.0）（J. Mol. Biol. 215: 403-10 (1990)）を用いて行うことができる。BLASTヌクレオチド検索は、本発明の核酸分

10

20

30

40

50

子に相同性のあるヌクレオチド配列を得るために、NBLASTプログラムを用い、score=100、wordlength=12で行うことができる。BLASTタンパク質検索は、本発明のタンパク質に相同性のあるアミノ酸配列を得るために、XBLASTプログラムを用い、score=50、wordlength=3で行うことができる。比較目的のギャップアラインメントを得るために、AltschulらのGapped BLAST (Nucleic Acids Res. 25 (17) : 3389-3402 (1997))を用いることができる。BLAST及びGapped BLASTプログラムを用いる際には、各プログラム（例えば、XBLASTやNBLAST）の既定のパラメータを用いることができる。

【0055】

本発明のペプチドのうちの一つを含むタンパク質において、プロセッシングを受ける前の全長形態と同様に、成熟プロセス形態は、本発明の薬剤一代謝酵素ペプチドのうちの1つに対して完全な配列同一性を有し、本発明の薬剤一代謝酵素ペプチドと同じ遺伝子位置によりコード化されているものとして、容易に同定することができる。本発明の新規な薬剤一代謝タンパク質をコード化している遺伝子は、（図3に示されるように）ヒト染色体1上にマップされるゲノム成分上に位置しており、これはSTS及びBACマップデータのような多系統の証拠によって支持されている。

【0056】

薬剤一代謝酵素ペプチドの対立遺伝子変異体は、薬剤一代謝酵素ペプチドの少なくとも一部に対して高度の（著しい）配列相同性／同一性を有するヒトタンパク質であることと同様に、本発明の薬剤一代謝酵素ペプチドと同じ遺伝子位置においてコード化されているものとして、容易に同定することができる。遺伝子位置は、対象となるヒトに対してマッピングされたゲノム配列のような、図3に示されるゲノム情報に基づいて容易に決定することができる。本発明の新規な薬剤一代謝タンパク質をコード化している遺伝子は、（図3に示されるように）ヒト染色体1上にマップされるゲノム成分上に位置しており、これはSTS及びBACマップデータのような多系統の証拠によって支持されている。ここで用いられる場合、アミノ酸配列において、典型的には少なくとも約70～80%、80～90%、さらに典型的には少なくとも約90～95%、又はそれ以上の相同性を有する場合には、2つのタンパク質（又はタンパク質の一領域）は著しい相同性を有している。本発明によれば、著しい相同性を有するアミノ酸配列は、より詳細には以下に述べられるようなストリンジェントな条件の下で、薬剤一代謝酵素ペプチドをコード化する核酸分子とハイブリダイズする核酸配列によりコード化される。

【0057】

図3には、本発明の薬剤一代謝酵素タンパク質をコード化している遺伝子において見出されたSNP情報を示す。挿入／欠失変異体（“indels”）を含むSNPは、45の異なるヌクレオチド位置で確認された。これらのSNPにより起こるアミノ酸配列の変化は、ユニバーサル遺伝子コード、及び図2に示されるタンパク質配列を用いて容易に確認することができる。エキソン、イントロン、及びORFの外側のそれぞれのSNPの位置は、それぞれのSNPより得られるDNA位置、及びその性質より得られる開始／終止、エキソン及びイントロン座標を用いて容易に決定することができる。

【0058】

薬剤一代謝酵素ペプチドのパラログは、薬剤一代謝酵素ペプチドの少なくとも一部に対して、ある程度の著しい配列相同性／同一性を有し、ヒトの遺伝子によってコード化され、また同様の活性又は機能を有しているものとして、容易に同定することができる。アミノ酸配列が、与えられた領域又はドメインを通じて、典型的には少なくとも約60%以上、さらに典型的には約70%以上の相同性を有する場合、2つのタンパク質は典型的なパラログであると考えられる。このようなパラログは、より詳細には以下に述べられるような、モデレートからストリンジェントな条件の下で、薬剤一代謝酵素ペプチドをコード化する核酸分子とハイブリダイズする核酸配列によりコード化される。

【0059】

薬剤一代謝酵素ペプチドのオルトログは、薬剤一代謝酵素ペプチドの少なくとも一部に対してある程度の著しい配列相同性／同一性を有し、他の生体の遺伝子によってコード化されているものとして、容易に同定することができる。オルトログは、好適には、哺乳類、さらに好適には霊長類から単離され、ヒトの治療ターゲット及び治療薬剤の開発のために用いられる。このようなオルトログは、より詳細には以下に述べられるような、モデレートからストリンジェントな条件の下で、薬剤一代謝酵素ペプチドをコード化する核酸分子とハイブリダイズするような核酸配列によりコード化され、これはタンパク質を生成する2つの生体の関連性に依存している。

【0060】

本発明の薬剤一代謝酵素ペプチドの非天然の変異体は、組み換え技術を用いて容易に生成することができる。このような変異体には、薬剤一代謝酵素ペプチドのアミノ酸配列中における欠失、付加、置換によるものが含まれるが、これに限定されるものではない。例えば、置換の1種として、保存的アミノ酸置換が挙げられる。この置換は、薬剤一代謝酵素ペプチドにおける特定のアミノ酸が、同様の特性をもつ他のアミノ酸によって置換されるものである。保存的置換においては、脂肪族のアミノ酸 *Ala*、*Val*、*Leu*、*Ile* の中の一つが他の一つに置換、ヒドロキシル残基 *Ser* と *Thr* との交換、酸性残基 *Asp* と *Glu* との交換、アミド残基 *Asn* と *Gln* 間の置換、塩基性残基 *Lys* と *Arg* との交換、芳香族残基 *Phe* と *Tyr* との置換がある。表現型に現れないアミノ酸置換に関する指針については、*Bowie et al.*, *Science* 247:1306-1310 (1990) に述べられている。

【0061】

変異薬剤一代謝酵素ペプチドは、完全に機能しているか、又は、例えば基質結合能、基質リン酸化能、信号伝達媒介能等のような活性の一つ以上において機能が欠落していることがある。完全機能的な変異体には、典型的には、保存的な変異、又は致命的でない残基あるいは領域内での変異体のみが含まれる。図2には、タンパク質解析の結果が示されており、致命的なドメイン／領域を特定するのに使用することができる。機能性変異体には、機能が変化しない、又は著しい機能変化の無い類似アミノ酸の置換も含まれる。他方、このような置換は、機能に対してある程度プラス又はマイナスの影響を及ぼすことがある。

【0062】

非機能性変異体には、典型的には、一つ以上の非保存的なアミノ酸の置換、欠失、導入、反転、切断、又は致命的な残基あるいは領域での置換、欠失、導入、反転が含まれる。

【0063】

機能において必須のアミノ酸は、例えば、特定部位の突然変異誘発、又はアラニンスキャニング突然変異誘発 (*Cunningham et al.*, *Science* 244:1081-1085 (1989)) 等の当該分野において既知の方法により、特に図2に示す結果を用いて確認することができる。後者の手順では、分子内の全ての残基において、単独のアラニン突然変異を導入する。この結果生じた変異体分子は、その後、薬剤一代謝酵素活性のような生物活性、又は *in vitro* 増殖活性分析のようなアッセイのために試験される。結合対象／基質結合にとって重要な部位は、結晶化、核磁気共鳴、光学親和性ラベル等の構造解析によって決定される (*Smith et al.*, *J. Mol. Biol.* 224:899-904 (1992); *de Vos et al.*, *Science* 255:306-312 (1992))。

【0064】

本発明は薬剤一代謝酵素ペプチドのフラグメントを提供し、さらにこれに加えて、該フラグメント、特に図2に特定された残基を含むフラグメントを含む、及びから成るタンパク質及びペプチドを提供するものである。しかしながら、本発明に関連するフラグメントは、本発明より以前に公開されているフラグメントを含むものとは見なされない。

【0065】

フラグメントは、ここで用いられる場合、薬剤一代謝酵素ペプチドに隣接するアミノ酸残基を、少なくとも8、10、12、14、16又はそれ以上含んでいる。このようなフラ

10

20

30

40

50

グメントは、1以上の薬剤一代謝酵素ペプチドの生物活性を保持する能力によって選択されるか、あるいは例えば、基質との結合又は抗原としての作用等の機能を果たす能力によって選択され得る。特に重要なフラグメントは生物活性フラグメントであり、これは例えば、8又はそれ以上のアミノ酸のペプチドである。このようなフラグメントは、典型的には、例えば活性部位、膜貫通ドメイン、又は基質結合ドメインのような、薬剤一代謝酵素ペプチドのドメイン又はモチーフを含んでいる。さらに、可能なフラグメントとしては、ドメイン又はモチーフ含有フラグメント、溶解性ペプチドフラグメント、抗原性構造含有フラグメントを含むが、これに限定されるものではない。予想されるドメイン及び機能性部位は、当業者にとって容易に入手可能な公知のコンピュータプログラム（例えばPROSITE分析）により、容易に確認することができる。このような分析結果の1つを図2

10

【0066】

ポリペプチドは、一般に20天然アミノ酸と呼ばれている20種のアミノ酸以外のアミノ酸をしばしば含むことがある。さらに、末端アミノ酸を含む多くのアミノ酸は、プロセッシング及び翻訳後修飾等の自然の過程、又は当該分野において公知の化学修飾技術によって修飾され得る。薬剤一代謝酵素ペプチドにおいて自然に生じる一般的な修飾については、基本的なテキスト、詳細な研究論文及び文献に記述されており、これは当業者においては周知である（これらの特性のいくつかは図2において確認される）。

【0067】

既知の修飾としては、アセチル化、アシル化、ADPリボシル化、アミド化、フラビンの共有結合付加、ヘム部分の共有結合付加、ヌクレオチド又はヌクレオチド誘導体の共有結合付加、脂質又は脂質誘導体の共有結合付加、ホスファチジルイノシトールの共有結合付加、架橋結合、環化、ジスルフィド結合形成、脱メチル化、共有結合架橋の形成、シスチンの形成、ピログルタミン酸塩の形成、ホルミル化、 γ -カルボキシル化、グリコシル化、GPIアンカー形成、ヒドロキシル化、ヨウ素化、メチル化、ミリストイル化、酸化、タンパク質分解プロセッシング、リン酸化、プレニル化、ラセミ化、セレノイル化、硫酸化、アルギニル化などのタンパク質へのアミノ酸の転移RNA媒介付加、ユビキチン化を含むが、修飾はこれに限定されるものではない。

20

【0068】

このような修飾は当業者には周知であり、科学文献において非常に詳細に記述されてきた。グリコシル化、脂質付加、硫酸化、グルタミン酸残基の γ -カルボキシル化、ヒドロキシル化、ADPリボシル化など、特に一般的な修飾は、Proteins - Structure and Molecular Properties, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993)のような、殆どの基本テキストにおいて記述されている。この点に関する詳細な文献としては、Wold, F., Posttranslational Covalent Modification of Proteins, B. C. Johnson, Ed., Academic Press, New York 1-12 (1983); Seifter et al. (Meth. Enzymol. 182: 626-646 (1990)) and Rattan et al. (Ann. N. Y. Acad. Sci. 663: 48-62 (1992))のような多くの文献を利用することができる。

30

40

【0069】

したがって、本発明の薬剤一代謝酵素ペプチドは、置換されたアミノ酸残基が遺伝子コードによってコード化されていない、置換基が含まれている、成熟薬剤一代謝酵素ペプチドが、例えば薬剤一代謝酵素ペプチドの半減期を長くする化合物のような他の化合物（例えば、ポリエチレングリコール）と融合しているか、又は付加アミノ酸が、例えば成熟薬剤一代謝酵素ペプチドの主、副配列、又は精製配列、あるいは前タンパク質配列のような、成熟薬剤一代謝酵素ペプチドと融合しているといった、誘導体又は類似体をも包含するものである。

50

【0070】

タンパク質／ペプチドの使用

本発明のタンパク質は、図面及び背景技術に示される機能情報に関連した、実質的及び特異的なアッセイにおいて、例えば、抗体を高める、又は他の免疫反応を誘発させるため、生物液中でのタンパク質（又はその結合対象、又はリガンド）レベルの定量のためのアッセイに用いる試薬（ラベル化試薬を含む）として、あるいは、対応するタンパク質を選択的に発現する組織のマーカー（組織の分化又は発達のある特定段階、あるいは疾患の状態）として使用することができる。タンパク質が、他のタンパク質と結合しているか又は結合し得る場合（例えば、薬剤－代謝酵素－効果器タンパク質相互作用、薬剤－代謝酵素－リガンド間相互作用）、このタンパク質を用いて結合相手を同定し、結合相互作用の阻害剤を同定するシステムを開発することができる。これらの一部又は全てを使用することにより、試薬グレード、又は商業製品としてのキットフォーマットへの開発が可能となる。

10

【0071】

上に列記した使用を実際に行う方法は、当業者にとって周知のことである。このような方法を開示している参考文献としては、“Molecular Cloning: A Laboratory Manual” 2nd ed, Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch and T. Maniatis eds., (1989) “Methods in Enzymology: Guide to Molecular Cloning Techniques”, Academic Press, Berger, S. L. and A. R. Kimmel eds., (1987) がある。

20

【0072】

本発明のペプチドの潜在的な用途は、主にタンパク質の起源に基づいており、同様にタンパク質の分類／作用に基づいている。例えば、ヒトから単離した薬剤－代謝酵素、及びそれらのヒト／哺乳類オルトログは、哺乳類の治療用医薬、例えば、ヒト用の医薬、特に、薬剤－代謝酵素を発現する細胞又は組織中での生物反応又は病理学的反応の変調に用いられる医薬を発見するためのターゲットとして有用である。図1の実験データによると、本発明の薬剤－代謝酵素タンパク質は、ヒトの胃、脳（乳児を含む）、子宮内膜腫瘍、前立腺、腎臓、副腎腫瘍、頭部／頸部、交感神経幹、胸部、及び肝細胞腫瘍で発現することが、仮想ノーザンブロット分析により示されている。PCRベースの組織スクリーニングパネルでは、脳で発現することも示されている。医薬製剤のうちの多くの割合のものが、薬剤－代謝酵素タンパク質、特にω-ヒドロキシラーゼ シトクロム P450 サブファミリーメンバーの活性を変調するものとして開発されている（背景技術参照）。背景技術及び図面に記載されている構造情報及び機能情報、特に図1の発現情報と組み合わせることによって、本発明の分子の特異的及び本質的な使用が提供される。図1の実験データによると、ヒトの胃、脳（乳児を含む）、子宮内膜腫瘍、前立腺、腎臓、副腎腫瘍、頭部／頸部、交感神経幹、胸部、及び肝細胞腫瘍で発現することが示されている。このような使用は、ここに示されている情報と、当業者において既知の情報、及びルーチン実験を用いて、容易に決定することができる。

30

【0073】

本発明のタンパク質（本発明以前に開示されている変異体及びフラグメントを含む）は、ω-ヒドロキシラーゼ シトクロム P450 サブファミリーに関連付けられる薬剤－代謝酵素に関連する生物学的アッセイに有用である。このようなアッセイは、特にこの薬剤－代謝酵素を発現する細胞及び組織において、本発明の一つが属する薬剤－代謝酵素サブファミリーに特異的な薬剤－代謝酵素関連症状の診断及び治療に有用な、公知の薬剤－代謝酵素の機能、活性、あるいは性質の何れかに関係している。図1の実験データによると、本発明の薬剤－代謝酵素タンパク質は、ヒトの胃、脳（乳児を含む）、子宮内膜腫瘍、前立腺、腎臓、副腎腫瘍、頭部／頸部、交感神経幹、胸部、及び肝細胞腫瘍で発現することが、仮想ノーザンブロット分析により示されている。PCRベースの組織スクリーニングパネルでは、脳で発現することも示されている。

40

50

【0074】

本発明のタンパク質は、細胞系、又は無細胞系における薬剤スクリーニングアッセイにおいても有用である (Hodgson, Bio/technology, 1992, Sept 10 (9): 973-80)。細胞系では、天然型、すなわち薬剤一代謝酵素を正常に発現する細胞であり、生体組織検査、又は細胞培地中で増殖することができる。図1の実験データによると、ヒトの胃、脳（乳児を含む）、子宮内膜腫瘍、前立腺、腎臓、副腎腫瘍、頭部／頸部、交感神経幹、胸部、及び肝細胞腫瘍で発現することが示されている。他の例では、細胞系アッセイは、薬剤一代謝酵素タンパク質を発現する組み換え宿主細胞に関係している。

【0075】

ポリペプチドは、自然状態又は変異型で薬剤一代謝酵素に関連する特定の疾患又は症状を引き起こすタンパク質の薬剤一代謝酵素活性を変調する化合物を同定するために用いることができる。本発明の薬剤一代謝酵素と、適切な変異体及びフラグメントの両者は、薬剤一代謝酵素への結合能力を持つ候補化合物をアッセイするための高効率スクリーンにおいて使用することができる。これらの化合物は、さらにこれらの薬剤一代謝酵素活性に対する影響を判定するために、機能性薬剤一代謝酵素に対してスクリーニングを行うことができる。さらにこれらの化合物は、動物又は無脊椎動物系における、活性／効果を判定するために試験することができる。化合物は、薬剤一代謝酵素を望ましい程度までに活性化（アゴニスト）又は不活性化（アンタゴニスト）するかどうか判定される。

【0076】

さらに、本発明のタンパク質は、薬剤一代謝酵素タンパク質と、薬剤一代謝酵素タンパク質と通常相互作用する分子、例えば、薬剤一代謝酵素タンパク質が通常相互作用する信号経路の基質又は構成成分との間での相互作用を促進又は阻害する能力を持つ化合物（例えば、他の薬剤一代謝酵素）をスクリーニングするのに用いることができる。このようなアッセイでは、一般的には、薬剤一代謝酵素タンパク質又はフラグメントが、ターゲット分子と相互作用し、タンパク質とターゲットとの複合物を検出するか、又はタンパク質のリン酸化、cAMP代謝回転、アデニル酸シクラーゼ活性化などの信号伝達に関連した効果のような、薬剤一代謝酵素タンパク質とターゲットとの間の相互作用の生化学的結果を検出することのできる条件で、薬剤一代謝酵素タンパク質と候補化合物とが結合される工程が含まれる。

【0077】

候補化合物としては、例えば、1) 最後部がIgの融合ペプチド、及びランダムペプチドライブラリを含む可溶性ペプチド（例えば、Lam et al., Nature 354: 82-84 (1991); Houghten et al., Nature 354: 84-86 (1991) 参照）、及びD-及び／又はL-型アミノ酸からできている組み合わせ化学誘導分子ライブラリを含むペプチド、2) ホスホペプチド（例えば、ランダム、又は部分的に変質したホスホペプチドライブラリ。例えば、Songyang et al., Cell 72: 767-778 (1993) 参照）、3) 抗体（例えば、ポリクローン抗体、モノクローン抗体、ヒト化抗体、抗イディオタイプ抗体、キメラ抗体、Fab、F(ab')₂、Fab発現ライブラリフラグメントを含む単鎖抗体、及び抗体のエピトープ結合フラグメント）、4) 小さな有機及び無機分子（例えば、組み合わせ及び天然生成物ライブラリから得られる分子）が含まれる。

【0078】

ある候補化合物は、基質結合と競合する受容体の可溶性フラグメントである。他の候補化合物には、変異薬剤一代謝酵素、又は薬剤一代謝酵素機能に影響を及ぼす変異体を含む適切なフラグメントがあり、このために基質との競合が起こる。したがって、例えば、高い親和性を有するか、又はフラグメントが基質と結合し解離しないような、基質と競合するフラグメントが本発明に含まれる。

【0079】

薬剤一代謝酵素により媒介される、生物学的又は生化学的な機能は、何れもエンドポイン

10

20

30

40

50

トアッセイとして用いることができる。これらは、ここに記載されている全ての生化学的又は生化学的／生物学的挙動を含み、ここに引用される文献にはこれらのエンドポイントアッセイのターゲットが折り込まれており、また、他の機能については、当業者において公知であるか、又は図面、特に図2の情報を用いて、容易に確認することができる。特に、薬剤一代謝酵素を発現する細胞又は組織の生物学的機能についてのアッセイを行うことができる。図1の実験データによると、本発明の薬剤一代謝酵素タンパク質は、ヒトの胃、脳（乳児を含む）、子宮内膜腫瘍、前立腺、腎臓、副腎腫瘍、頭部／頸部、交感神経幹、胸部、及び肝細胞腫瘍で発現することが、仮想ノーザンブロット分析により示されている。PCRベースの組織スクリーニングパネルでは、脳で発現することも示されている。

【0080】

結合及び／又は活性化化合物は、キメラ薬剤一代謝酵素タンパク質を用いることによりスクリーニングを行うこともでき、アミノ末端細胞外ドメイン又はその一部、あるいは、7つの膜貫通セグメントの何れか又は細胞内又は細胞外ループの何れか、及びカルボキシ末端細胞内ドメインのような膜貫通ドメイン全体あるいはサブリジョン又はその一部は、異種ドメイン又はサブリジョンにより置換され得る。例えば、基質結合領域は、異なる基質と相互作用するものとして用いられることができ、この場合、天然の薬剤一代謝酵素によって認識される。したがって、異なる一組の信号伝達成分を、活性化のエンドポイントアッセイとして利用することができる。これにより、薬剤一代謝酵素の由来する特定の宿主細胞以外のものにおいてアッセイを行うことが可能となる。

【0081】

本発明の薬剤一代謝酵素ポリペプチドはまた、薬剤一代謝酵素と相互作用する化合物（例えば、結合相手及び／又はリガンド）を発見するため設計される方法である競争結合アッセイにも有用である。このために、化合物がポリペプチドと結合又は相互作用できる条件下で、化合物を薬剤一代謝酵素ポリペプチドと接触させる。可溶性薬剤一代謝酵素ポリペプチドもまた混合物中に加えられる。テスト化合物が可溶性薬剤一代謝酵素ポリペプチドと相互作用する場合、薬剤一代謝酵素ターゲットから形成される複合体の量、又は活性は減少する。このタイプのアッセイは、特に薬剤一代謝酵素の特定領域と相互作用する化合物を模索する場合に有用である。このため、ターゲットとなる薬剤一代謝酵素領域と競合する可溶性ポリペプチドは、対象となる領域に対応したペプチド配列を含むように設計されている。

【0082】

無細胞系薬剤のスクリーニングアッセイを行うためには、タンパク質の一方又は両方の非複合化形態からの複合化形態の分離を効率化し、アッセイを自動化するために、薬剤一代謝酵素タンパク質又はフラグメント、そのターゲット分子の何れかを固定化するのが望ましい場合がある。

【0083】

薬剤スクリーニングアッセイにおいては、マトリックスにタンパク質を固定化する技術を使用することができる。ある例では、タンパク質にマトリックスに結合することのできるドメインを付加した融合タンパク質を得ることができる。例えば、グルタチオンS-トランスフェラーゼ融合タンパク質は、グルタチオンセファローズビーズ（Sigma Chemical, St. Louis, MO）又はグルタチオン誘導マイクロタイタープレート上に吸着され、細胞溶解物（例えば、³⁵S-labeled）と候補化合物とが結合され、複合体形成条件（例えば、塩及びpHの生理学的条件）の下で混合物がインキュベートされる。インキュベートの後、非結合ラベルの除去のためにビーズは洗浄され、マトリックスを固定化して、放射性ラベルを直接、又は複合体を分離した後の上澄みを測定することにより定量される。あるいは、複合体はSDS-PAGEによりマトリックスから分離することができ、標準の電気泳動技術を用いることによって、ゲルからビーズフラクション中の薬剤一代謝酵素結合タンパク質のレベルを定量することができる。例えば、ポリペプチド又はそのターゲット分子のどちらかが、当業者に周知の技術を利用して、ビオチン及びストレプトアビジンの結合を用いて固定化される。あるいは、タンパク質と

10

20

30

40

50

反応し、タンパク質とターゲット分子との結合を妨げない抗体は、プレートのウェルに誘導化され、このタンパク質は抗体との結合によりそのウェルの中に捕らえられる。薬剤一代謝酵素結合タンパク質の組成物と候補化合物は、薬剤一代謝酵素タンパク質存在ウェル中でインキュベートされ、ウェルに捕らえられた複合体の量を測定する。このような複合体を検出する方法としては、GST固定複合体による前述の方法に加えて、薬剤一代謝酵素タンパク質ターゲット分子に反応性のある抗体、又は、薬剤一代謝酵素タンパク質に反応性がありターゲット分子と競合する抗体を用いた複合体の免疫検出法、及びターゲット分子と関連する酵素活性の検出に基づく酵素結合アッセイが含まれる。

【0084】

本発明の薬剤一代謝酵素のうちの1つを変調する薬剤は、上述の分析方法を単独あるいは組み合わせる用いることにより同定することができる。一般的には、最初と最後に細胞系又は無細胞系のシステムを用いて、動物又は他のモデルシステムにおける活性を確認することが好ましい。このようなモデルシステムは、当業者に周知であり、本記載において容易に用いることができる。

10

【0085】

これらの薬剤スクリーニングアッセイによって同定される薬剤一代謝酵素タンパク質活性のモジュレータは、薬剤一代謝酵素を発現する細胞又は組織に処置することにより、薬剤一代謝酵素経路により媒介される疾患を患う患者の治療に用いることができる。図1の実験データによると、ヒトの胃、脳（乳児を含む）、子宮内膜腫瘍、前立腺、腎臓、副腎腫瘍、頭部／頸部、交感神経幹、胸部、及び肝細胞腫瘍で発現することが示されている。これらの治療方法には、薬剤組成物中の薬剤一代謝酵素活性のモジュレータを患者の治療に必要な量投与する工程が含まれており、このモジュレータはここに記載されるようにして同定される。

20

【0086】

本発明の他の観点では、薬剤一代謝酵素と結合又は相互作用する、又は薬剤一代謝酵素活性に関連している他のタンパク質を同定するために、2-ハイブリッドアッセイ又は3-ハイブリッドアッセイ（U. S. Patent No. 5, 283, 317; Zervos et al. (1993) Cell 72: 223-232; Madura et al. (1993) J. Biol. Chem. 268: 12046-12054; Bartel et al. (1993) Biotechniques 14: 920-924; Iwabuchi et al. (1993) Oncogene 8: 1693-1696; and Brent WO 94/10300 参照）において、薬剤一代謝酵素タンパク質を「baitタンパク質」として使用することができる。このような薬剤一代謝酵素結合タンパク質は、例えば、下流因子としての薬剤一代謝酵素タンパク質、又は薬剤一代謝酵素ターゲットによる信号伝達に関与している。あるいは、このような薬剤一代謝酵素結合タンパク質は、薬剤一代謝酵素阻害剤であることがある。

30

【0087】

2-ハイブリッドシステムは、分離可能なDNA結合ドメイン及び活性化ドメインから成る、大部分の転写調節因子のモジュラー性に基づいている。簡単に言うと、このアッセイでは2つの異なるDNA構造を利用する。一方の構造においては、薬剤一代謝酵素タンパク質をコードしている遺伝子は、既知の転写調節因子（例えばGAL-4）のDNA結合ドメインをコード化している遺伝子に融合される。他方の構造においては、DNA配列ライブラリから得られ、未確認のタンパク質（「prey」又は「sample」）をコード化しているDNA配列が、既知の転写調節因子の活性化ドメインをコード化している遺伝子に融合される。「baitタンパク質」及び「preyタンパク質」が生体内で相互作用することができ、薬剤一代謝酵素依存複合体を形成する場合、転写調節因子のDNA結合ドメイン及び活性化ドメインは近接する。この近接により、転写調節因子に反応する転写調節部位に結合可能なリポーター遺伝子（例えばlacZ）の転写を行うことができる。リポーター遺伝子の発現は検出することが可能であり、機能的転写調節因子を含んでいる細胞コロニーを単離及び使用して、薬剤一代謝酵素タンパク質と相互作用するタンパ

40

50

ク質をコード化するクローン遺伝子を得ることができる。

【0088】

本発明はさらに、前述のスクリーニングアッセイによって同定される新規な薬剤に関する。したがって、ここに記載されるようにして同定された薬剤を適当な動物のモデルに使用することも本発明の範囲内である。例えば、本発明で同定された薬剤（例えば薬剤一代謝酵素変調薬剤、アンチセンス薬剤一代謝酵素核酸分子、薬剤一代謝酵素特異抗体、又は薬剤一代謝酵素結合対象）は、これらの薬剤の処方における有効性、毒性、副作用を判定するために、動物、又は他のモデルで用いることができる。あるいは、本発明で同定された薬剤は、このような薬剤の作用のメカニズムを決定するために、動物又は他のモデルで使われることができる。さらに、本発明は、ここに記載される治療のためのスクリーニング

10

【0089】

本発明の薬剤一代謝酵素タンパク質は、ペプチドにより媒介される疾患又は疾患素質の診断のためのターゲットを提供するのに有用である。したがって、本発明は、細胞、組織、又は生体中におけるタンパク質（又はコード化したmRNA）の存在、あるいはそのレベルを検出する方法を提供するものである。図1の実験データによると、ヒトの胃、脳（乳児を含む）、子宮内膜腫瘍、前立腺、腎臓、副腎腫瘍、頭部／頸部、交感神経幹、胸部、及び肝細胞腫瘍で発現することが示されている。この方法は、生物サンプルと、薬剤一代謝酵素タンパク質との相互作用能力を有する化合物との接触に関係しており、ここでは相互作用について検出することができる。このようなアッセイは、単検出形態、又は抗体チ

20

【0090】

サンプル中のタンパク質を検出する1つの薬剤は、タンパク質に選択的に結合することのできる抗体である。生物サンプルには、被験者から単離されるか、又は被験者の内部に存在する組織、細胞、生物液体が含まれる。

【0091】

本発明のペプチドは、ペプチド変異体を持つ患者のタンパク質の活性、疾患又は疾患素質の診断に用いるためのターゲット、特に現存するタンパク質ファミリー以外のものとして知られる活性、及び症状のターゲットを提供するものである。したがって、ペプチドは生物学的サンプルから単離されることができ、また、異常ペプチドを生じる遺伝子突然変異の存在についてアッセイを行うことができる。これは、アミノ酸置換、欠失、挿入、再配置（異常なスプライシング挙動の結果生じる）、及び不適当な翻訳後の修飾を含む。分析方法としては、変容電気泳動移動度、変容トリプシンペプチド消化、細胞系又は無細胞系アッセイによる変容薬剤一代謝酵素活性、基質又は抗体の変容結合パターン、変容等電点、直接アミノ酸配列、及びタンパク質の変異の検出に有用な他の公知のアッセイ技術を含む。このようなアッセイは、単検出形態、又は抗体チップアレイのような多検出形態で与えられる。

30

【0092】

ペプチドの *In vitro* 検出技術としては、酵素結合免疫吸着剤アッセイ（ELISA）、ウェスタンブロット、抗体、又はタンパク質結合剤のような検出試薬を用いた免疫沈降、免疫蛍光検査法を含む。あるいは、ラベルされた抗ペプチド抗体、又は他のタイプの検出試薬を被験者に導入することにより、被験者の *in vivo* 検出を行うことができる。例えば、被験者中のこのマーカーの存在及び位置を標準のイメージング技術によって検出することができる放射性マーカーを、抗体にラベルすることができる。被験者において発現されたペプチドの対立変異体を検出する方法、及びサンプル中のペプチドのフラグメントを検出する方法は、特に有用である。

40

【0093】

ペプチドはまた、ゲノム薬理学分析においても有用である。ゲノム薬理学では、薬剤の変化の傾向と、影響を受けたヒトの異常挙動に従って、薬剤に対しての応答における著しい遺伝的変異について臨床的に取り扱う。例えば、Eichelbaum, M. (C

50

n. Exp. Pharmacol. Physiol. 23 (10-11) : 983-985 (1996))、及びLinder, M. W. (Clin. Chem. 43 (2) : 254-266 (1997))を参照されたい。これらの変異の臨床的な結果、個人の代謝変異の結果として、ある個人に対しては治療薬剤が重い毒性となり、又はある個人に対しては治療の失敗に終わる。このように、個人の遺伝子型は、体内で治療化合物を作用させる方法、又は体が化合物を代謝する方法を決定することができる。さらに、酵素を代謝させる薬剤の活性は、薬剤の作用の強度と期間に影響する。このように、個人のゲノム薬理学は、個人の遺伝子型に基づいた予防、又は治療的な処置において、効果的な化合物、又はこのような化合物の効果的な投与量の選択を可能とする。遺伝子多形性の発見により、酵素代謝性の薬剤において、患者が期待される薬効を得られない、不自然な薬効を示す、又は標準の投薬量から重大な毒性を被るといったことの理由を説明することができる。多形性は、強い代謝系の表現型と弱い代謝系の表現型で表されることができる。したがって、遺伝子の多形性は、ある集団の薬剤一代謝酵素機能の1つ以上が他の集団のそれと異なるような、薬剤一代謝酵素タンパク質の対立タンパク質変異に至るかもしれない。このように、ペプチドは治療法に影響する遺伝子の素因を確認するためのターゲットとなり得る。このため、リガンドベースの治療において、多形性は、末端アミノ基の細胞外ドメイン及び／又は他の基質結合領域における基質結合活性、及び薬剤一代謝酵素活性が、より高い、又はより低いことを引き起こし得る。したがって、多形性を含む集団においては、治療効果を最大にするように、基質投薬量は必然的に修正される。遺伝子型に代わるものとしては、特定の多形性のペプチドを同定することができる。

10

20

【0094】

ペプチドはまた、タンパク質の発現がない、発現が不適當、又は発現が不必要であることによって特徴づけられる障害の治療に有用である。図1の実験データによると、ヒトの胃、脳（乳児を含む）、子宮内膜腫瘍、前立腺、腎臓、副腎腫瘍、頭部／頸部、交感神経幹、胸部、及び肝細胞腫瘍で発現することが示されている。したがって、治療方法としては、薬剤一代謝酵素タンパク質又はフラグメントの使用が含まれる。

【0095】

抗体

本発明は、本発明のペプチド、このようなペプチドを含むタンパク質、それらの変異体及びフラグメントの1つに選択的に結合する抗体を提供するものである。ここで用いられている場合、抗体がターゲットペプチドと結合し、無関係なタンパク質と強く結合しないような場合、抗体は選択的にターゲットペプチドと結合している。抗体がターゲットペプチドと実質的に相同性の無い他のタンパク質と結合していても、そのペプチドが抗体のターゲットとなるペプチドに対してフラグメント又はドメインにおける相同性を有している限り、抗体は選択的にペプチドを結びつけると考えられる。この場合、ペプチドに結合している抗体は、ある程度の交差反応性を持つにも関わらず、選択的であると理解される。

30

【0096】

ここで用いられる場合、抗体は当該分野で認められているものと同じ用語で定義され、これらは、哺乳類生体により生成され、抗原の攻撃に対して応答するマルチサブユニットタンパク質である。本発明の抗体には、ポリクローナル抗体、モノクローナル抗体、及びF a b又はF (a b')₂、及びF vのような抗体のフラグメントが含まれるが、これに限定されるものではない。

40

【0097】

ターゲットペプチドを得るための、抗体の生成及び／又は同定について、多くの方法が知られている。このような方法のいくつかは、Harlow, Antibodies, Cold Spring Harbor Press, (1989)に記載されている。

【0098】

一般に、抗体を生成するためには、単離ペプチドを免疫原として用い、例えばラット、ウサギ又はマウスのような哺乳類生体に投与される。全長タンパク質、抗原性ペプチドフラグメント又は融合タンパク質が用いられる。特に重要なフラグメントは、図2に特定され

50

ているような機能性ドメインをカバーするものであり、また、タンパク質アラインメント法を使用し、図面に示されているようにして、容易に確認することができるファミリーと配列相同性又は相違性を持つドメインである。

【0099】

抗体は、薬剤一代謝酵素タンパク質の領域、又は単離されたフラグメントから好適に調製される。抗体は、ここに記載されるペプチドの何れの領域からも調製することができる。しかしながら、好適な領域としては、機能／活性及び／又は薬剤一代謝酵素結合対象相互作用に関係している領域が含まれる。図2は特に重要な領域を特定するのに用いることができ、配列アラインメントは保持された特異な配列フラグメントを特定するのに用いることができる。

10

【0100】

抗原性フラグメントは、一般的に、少なくとも8つの隣接するアミノ酸残基を含んでいる。しかしながら、抗原性ペプチドは、少なくとも10、12、14、16以上のアミノ酸残基を含むことができる。このようなフラグメントは、例えば、タンパク質の表面上に位置する領域、例えば、親水性領域に対応するフラグメントのような物理的な性質、あるいは配列の特異性（図2参照）に基づいて選択することができる。

【0101】

本発明の抗体の検出は、検出可能物質と抗体とのカップリング（すなわち、物理的な結合）によって容易にすることができる。検出可能物質の例としては、例えば、種々の酵素、補欠分子族、蛍光性物質、発光性物質、生物発光性物質、及び放射性物質が含まれる。好適な酵素の例としては、セイヨウワサビペルオキシダーゼ、アルカリホスファターゼ、 β -ガラクトシダーゼ、又はアセチルコリンエステラーゼが含まれ、好適な補欠分子族の例としては、ストレプトアビジン／ビオチン、アビジン／ビオチンが含まれ、好適な蛍光性物質の例としては、ウンベリフェロン、フルオレセイン、フルオレセイン・イソチオシアン酸塩、ローダミン、ジクロロトリアジニルアミンフルオレセイン、ダンシルクロライド、又はフィコエリトリンが含まれ、発光性物質の例としては、ルミノールが含まれ、生物発光性物質の例としては、ルシフェラーゼ、ルシフェリン及びエクオリンが含まれ、そして、適切な放射性物質の例としては、 ^{125}I 、 ^{131}I 、 ^{35}S 、又は ^3H が含まれる。

20

【0102】

抗体の使用

抗体は、アフィニティークロマトグラフィ、又は免疫沈降のような標準の技術によって、本発明のタンパク質の1つを単離するために用いることができる。抗体は、細胞からの天然タンパク質、及び宿主細胞に表現される組換えによって生成されたタンパク質の精製を容易にすることができる。さらに、このような抗体は、組織や生体内のタンパク質の発現パターンを決定するため、細胞や組織内における本発明のタンパク質の存在の検出に、通常の開発段階を経ずに用いることができる。図1の実験データによると、本発明の薬剤一代謝酵素タンパク質は、ヒトの胃、脳（乳児を含む）、子宮内膜腫瘍、前立腺、腎臓、副腎腫瘍、頭部／頸部、交感神経幹、胸部、及び肝細胞腫瘍で発現することが、仮想ノーザンブロット分析により示されている。PCRベースの組織スクリーニングパネルでは、脳で発現することも示されている。さらに、このような抗体は、発現の量及びパターンを評価することによって、*in situ*、*in vitro*、細胞溶解物中、及び上澄み中でのタンパク質の検出に用いることができる。また、このような抗体は、生物学的症状の発展又は進行間における、異常な組織分布、又は異常な発現を評価するのに用いることができる。全長タンパク質の循環フラグメントにおける抗体検出は、代謝回転を確認するのに用いることができる。

30

40

【0103】

さらに、抗体は、タンパク質機能に関連した疾患の活発な段階、又は該疾患素因を持つ個人といった、疾患の症状発現を評価するのに用いることができる。障害が不適当な組織分布、発展性の発現、タンパク質の発現レベル、又は発現／進行状態に起因する場合、抗体

50

を通常のタンパク質に対して調製することができる。図1の実験データによると、ヒトの胃、脳（乳児を含む）、子宮内膜腫瘍、前立腺、腎臓、副腎腫瘍、頭部／頸部、交感神経幹、胸部、及び肝細胞腫瘍で発現することが示されている。障害がタンパク質の特定の変異により特徴づけられる場合、この変異タンパク質に特異的な抗体を、特定の変異タンパク質の存在のアッセイを行うために用いることができる。

【0104】

抗体はまた、生体内の各種組織における、細胞の正常又は異常な細胞小器官の位置を評価するのに用いることができる。図1の実験データによると、ヒトの胃、脳（乳児を含む）、子宮内膜腫瘍、前立腺、腎臓、副腎腫瘍、頭部／頸部、交感神経幹、胸部、及び肝細胞腫瘍で発現することが示されている。診断としての使用は、遺伝子のテストだけでなく、治療法をモニターすることにも適用することができる。したがって、治療が最終的に、発現レベル、又は異常配列及び異常組織配布の存在、又は発展性の発現を修正することを目指すものである場合、タンパク質又は関連するフラグメントに直接的に対する抗体を、治療の有効性をモニターするために用いることができる。

10

【0105】

さらに、抗体はゲノム薬理分析において有用である。したがって、多形タンパク質に対して調製される抗体は、治療法の修正を必要とする個人を特定するために用いることができる。抗体はまた、電気泳動、等電点、トリプシンペプチド消化、当該分野において周知な他の物理的なアッセイによって分析される、異常タンパク質の免疫学的マーカーのような診断上のツールとしても有用である。

20

【0106】

抗体はまた、組織型の分類にも有用である。図1の実験データによると、ヒトの胃、脳（乳児を含む）、子宮内膜腫瘍、前立腺、腎臓、副腎腫瘍、頭部／頸部、交感神経幹、胸部、及び肝細胞腫瘍で発現することが示されている。このように、特定のタンパク質が特定の組織中の発現と相関していた場合、このタンパク質に特異な抗体を、組織型を同定するために用いることができる。

【0107】

抗体はまた、タンパク質機能を阻害するのに有用であり、例えば、基質やタンパク質結合相手のような結合対象への薬剤一代謝酵素ペプチドの結合をブロックする。これらの使用は、タンパク質の機能阻害に関連する治療法における、治療環境において適用されることができる。抗体は、例えば、結合をブロックすることにより、ペプチド活性の変調（アゴナイズ又はアンタゴナイズ）を阻害することに用いることができる。抗体は、機能のために必要な部位を含む特定のフラグメントに対して、又は細胞又は細胞膜と関係している完全タンパク質に対して調製される。図2に、本発明のタンパク質に関する構造情報を示す。

30

【0108】

本発明は、生物学的サンプル中のタンパク質の存在を検出するために抗体を用いたキットを包含する。キットには、ラベル化された、又はラベル化可能な抗体と、生物学的サンプル中でタンパク質を検出するための化合物又は試薬を含み、；サンプル中のタンパク質量を決定する手段；標準サンプルのタンパク質量と比較する手段；及び使用の説明、とを含む。このようなキットは、単一のタンパク質、又はエピトープを検出するために提供されるか、又は抗体検出アレイのように、多数のエピトープのうちの1つを検出するように設定されることができる。アレイとしては、核酸アレイが後述され、また抗体アレイのための同様の方法も開発されている。

40

【0109】

核酸分子

本発明は、さらに本発明の薬剤一代謝酵素ペプチド又はタンパク質をコード化する単離核酸分子（cDNA、転写及びゲノム配列）を提供するものである。このような核酸分子は、本発明の薬剤一代謝酵素ペプチドの1つ、又はこれらの対立変異体、又はこれらのオルトログ又はパラログをコード化する核酸分子から成る、実質的に成る、又は含んでいる。

50

【0110】

ここで用いられる場合、「単離」核酸分子とは、核酸の天然起源における他の核酸の存在から分離されたものである。好ましくは、「単離」核酸は、その核酸の由来となる生物のゲノムDNAにおいて、核酸のフランキング配列（すなわち、5'、及び3'末端に位置する配列）は含まない。しかしながら、例えば、約5KB、4KB、3KB、2KB又は特に1KB以上又はこれ以下、特に配列によりコード化された隣接ペプチドのように、いくつかのフランキングヌクレオチド配列があるが、同一の遺伝子の配列によりコード化されたペプチドは、ゲノム配列中のイントロンにより分離されている。重要な点は、核酸が、ここに記載されているような特定の操作、例えば、組み換え発現、プローブやプライマーの調製、核酸配列のための他の使用等に取り扱うことができるように、遠くの重要なフランキング配列から分離されているということである。

10

【0111】

さらに、例えば、転写／cDNA分子のような「単離」核酸分子は、組み換え技術により製造される場合には、他の細胞物質、又は培地、また化学的に合成される場合には前駆体又は他の化学物質を実質的に含まない。しかしながら、この核酸分子は、他のコード化又は調整配列に融合されることができ、これは単離されたものとして考えられる。

【0112】

例えば、ベクターに含まれる組み換えDNA分子は、単離されたものとして考えられる。さらに、単離したDNA分子の例としては、非相同的な宿主細胞に保持された組み換えDNA分子、又は溶液中で精製（部分的又は実質的に）されたDNA分子が含まれる。単離されたRNA分子としては、本発明の単離したDNA分子の、*in vivo*又は*in vitro* RNA転写物が含まれる。本発明により単離された核酸分子としてはさらに、合成的に製造された分子が含まれる。

20

【0113】

したがって、本発明は、図1又は3 [SEQ ID NO. 1 (転写配列)、及びSEQ ID NO. 3 (ゲノム配列)] において示されるヌクレオチド配列から成る核酸分子、又は図2 (SEQ ID NO. 2) に示されるタンパク質をコード化する核酸分子を提供するものである。ヌクレオチド配列がこの核酸分子の完全なヌクレオチド配列であるとき、核酸分子はヌクレオチド配列から成る。

【0114】

本発明はさらに、図1又は3 [SEQ ID NO. 1 (転写配列)、及びSEQ ID NO. 3 (ゲノム配列)] において示されるヌクレオチド配列から実質的に成る核酸分子、又は図2 (SEQ ID NO. 2) に示されるタンパク質をコード化する核酸分子を提供するものである。核酸分子において、最終的にこのようなヌクレオチド配列がごくわずかの付加核酸残基とともに存在するとき、核酸分子はヌクレオチド配列から実質的に成る。

30

【0115】

本発明はさらに、図1又は3 [SEQ ID NO. 1 (転写配列) 及びSEQ ID NO. 3 (ゲノム配列)] において示されるヌクレオチド配列を含む核酸分子、又は図2 (SEQ ID NO. 2) に示されるタンパク質をコード化する核酸分子を提供するものである。核酸分子の最終的な核酸配列の少なくとも一部がこの核酸配列である場合、核酸分子はヌクレオチド配列を含む。これによると、核酸分子は、そのヌクレオチド配列だけであるか、又は付加的な核酸残基、例えば、自然に関する核酸配列、又は非相同的な核酸配列を有することもできる。このような核酸分子は、ごくわずかの付加的なヌクレオチドを有するか、又は数百又はそれ以上の付加的なヌクレオチドを含むこともできる。これらの種々のタイプの核酸分子を容易に生成／単離する方法については、以下に簡単に述べる。

40

【0116】

図1及び3には、コード及び非コードの配列の両者が提供される。本発明のヒトゲノム配列 (図3)、及びcDNA／転写配列 (図1) より、図中の核酸分子は、ゲノムイントロン配列、5'と3'の非コード配列、遺伝子調節領域、及び非コード遺伝子間配列を含んでいる。一般に、図1と図3の両者に記載されているこのような配列の特徴は、当該分野

50

において公知の計算ツールを用いて容易に特定することができる。以下で議論されるように、いくつかの非コーディング部位、特にプロモーターのような遺伝子の調節因子は、例えば、非相同的な遺伝子発現の制御、遺伝子活性を変調する化合物同定のターゲット等の種々の目的にとって有用であり、ここに提供されるゲノム配列のフラグメントとして特別にクレームされている。

【0117】

単離した核酸分子は、成熟したタンパク質に加えて付加的アミノ末端あるいはカルボキシ末端アミノ酸、又は成熟ペプチド内のアミノ酸（例えば、成熟形態が1以上のペプチド鎖を有する場合）をコード化することができる。このような配列は、前駆体から成熟した形態へのタンパク質のプロセッシングにおいて、タンパク質搬送の促進、タンパク質半減期の延長あるいは短縮、又はタンパク質のアッセイ又は製造の際の操作の効率化、又は他の事象における役割を果たし得る。一般に、*in situ*の場合、付加アミノ酸は細胞酵素によって成熟したタンパク質へとプロセッシングされる。

10

【0118】

上述したように、単離した核酸分子は、単独で薬剤一代謝酵素ペプチドをコード化している配列、成熟したペプチドをコード化している配列、そして、主、又は副配列（例えば、*pre-pro*、*pro-protein*配列）のような付加的なコード配列を含むが、これに限定されるものではなく、付加的なコード配列に加えて、付加的な非コード配列、例えば、イントロンと非コード5'及び3'配列のような、転写されるものの翻訳はされずに、転写、mRNAプロセッシング（スプライシング及びポリアダニル化を含む）、リボソームの結合、及びmRNAの安定性の役割を果たすものを、含んでも含まなくても良い。加えて、核酸分子は、例えば、精製を容易にするようなペプチドをコード化した配列のマーカーと融合されることもできる。

20

【0119】

単離した核酸分子は、mRNAのようなRNAの形態、あるいはクローニング、化学合成技術又はその組み合わせによって生成するcDNA及びゲノムDNAを含むDNAの形態をとり得る。核酸、特にDNAは、二重鎖、又は単鎖であり得る。単鎖の核酸は、コード鎖（センス鎖）、又は非コード鎖（アンチセンス鎖）であり得る。

【0120】

本発明はさらに、本発明のペプチドのフラグメントをコード化する核酸分子と同様に、上記したような本発明の薬剤一代謝酵素タンパク質の明らかな変異体をコード化する核酸分子を提供するものである。このような核酸分子は、対立変異体（同一位置）、パラログ（異なる位置）とオルトログ（異なる生体）のように自然に発生するか、あるいは組み換えDNA法又は化学合成によって生成され得る。このような非自然に発生する変異体は、核酸分子、細胞又は生体に適用されるものを含む突然変異生成技術によって生成され得る。したがって、上述したように、変異体にはヌクレオチドの置換、欠失、倒置、挿入が含まれる。変異は、コード、非コード領域のどちらか、又は両方で起こることができる。変異は、保持及び非保持アミノ酸置換の両方で生じることができる。

30

【0121】

本発明はさらに、図1及び3に示される核酸分子の非コードのフラグメントを提供するものである。好適な非コードのフラグメントとしては、プロモーター配列、エンハンサ配列、遺伝子変調配列、遺伝子終止配列が含まれるが、これに限定されるものではない。このようなフラグメントは、非相同的な遺伝子発現の制御、及び遺伝子変調薬剤の同定を行うためのスクリーニングの開発において有用である。プロモーターは、図3のゲノム配列における5'からATG開始部位において容易に確認される。

40

【0122】

フラグメントは、12又はそれ以上の隣接するヌクレオチド配列を含む。さらに、フラグメントは少なくとも30、40、50、100、250、又は500のヌクレオチド長であり得る。フラグメントの長さは使用目的に基づく。例えば、フラグメントは、ペプチドのエピトープ結果領域をコード化することができるか、又はDNAプローブ及びDNAプ

50

ライマーとして有用である。このようなフラグメントは、オリゴヌクレオチドプローブを合成するための既知のヌクレオチド配列を用いて単離することができる。ラベル化されたプローブは、コード化領域と対応する核酸を単離するため、cDNAライブラリ、ゲノムDNAライブラリ、又はmRNAのスクリーニングに用いることができる。さらに、プライマーは、遺伝子の特定領域をクローンするためのPCR反応に用いることができる。

【0123】

プローブ／プライマーは、典型的には、実質的に精製されたオリゴヌクレオチド又はオリゴヌクレオチドペアを含む。オリゴヌクレオチドは、典型的には、少なくとも12、20、25、40、50又はそれ以上の連続的ヌクレオチドに、ストリンジェントな条件下でハイブリダイズされたヌクレオチド配列領域を含む。

10

【0124】

オルトログ、ホモログ、及び対立変異体は、当該技術分野において周知の方法を用いて同定することができる。ペプチドの項で述べたように、これらの変異体は、ペプチドをコード化するヌクレオチド配列を含み、図面に示されるヌクレオチド配列、又はこの配列のフラグメントに対して、典型的には、60-70%、70-80%、80-90%、より典型的には、少なくとも90-95%以上の相同性を有するものである。このような核酸分子は、図面に示されるヌクレオチド配列、又はこの配列のフラグメントに対して、モデルートな条件からストリンジェントな条件の下でハイブリダイズが可能なものとして、容易に同定することができる。対立変異体は、コード化している遺伝子の遺伝子位置により、容易に決定されることができる。本発明の新規な薬剤-代謝タンパク質をコード化している遺伝子は、(図3に示されるように) ヒト染色体1上にマップされるゲノム成分上に位置しており、これはSTS及びBACマップデータのような多系統の証拠によって支持されている。

20

【0125】

図3には、本発明の薬剤-代謝酵素タンパク質をコード化している遺伝子において見出されたSNP情報を示す。挿入／欠失変異体("indels")を含むSNPは、45の異なるヌクレオチド位置で確認された。これらのSNPにより起こるアミノ酸配列の変化は、ユニバーサル遺伝子コード、及び図2に示されるタンパク質配列を用いて容易に確認することができる。エキソン、イントロン、及びORFの外側のそれぞれのSNPの位置は、それぞれのSNPより得られるDNA位置、及びその性質より得られる開始／終止、エキソン及びイントロン座標を用いて容易に決定することができる。

30

【0126】

「ストリンジェントな条件下でハイブリダイズする」という用語は、ここで用いられる場合、互いにハイブリダイズして残存する、ペプチドをコード化しているヌクレオチド配列が、互いに少なくとも60-70%の相同性を有する程度にハイブリダイズ、及び洗浄が行われる条件を意味している。この条件では、ハイブリダイズして残る配列が、少なくとも約60%、少なくとも約70%、少なくとも約80%、又はそれ以上となる。このようなストリンジェントな条件は、当業者においては周知であり、Current Protocols in Molecular Biology, John Wiley & Sons, N. Y. (1989), 6.3.1-6.3.6に記載されている。ストリンジェントなハイブリダイズ条件の1つの例では、6X塩化ナトリウム／クエン酸ナトリウム(SSC)中、約45℃でハイブリダイズし、その後、0.2XSSC0.1%SDS中、50-65℃で洗浄する。低ストリンジェンシーのモデルートなハイブリダイズ条件の例は、当業者において周知である。

40

【0127】

核酸分子の使用

本発明の核酸分子は、プローブ、プライマー、化学合成中間体、及び生物学アッセイにおいて有用である。核酸分子は、図2に示されているペプチドをコード化する全長cDNA及びゲノムクローンを単離するため、及び図2に示すペプチドと同一又は関連したペプチドを生成する変異体(対立遺伝子、オルトログ等)に対応するゲノムクローンを単離する

50

ための、メッセンジャーRNA、転写／cDNA、及びゲノムDNAのハイブリダイゼーションプローブとして有用である。図3に示されるように、SNPは45の異なるヌクレオチド位置で確認された。

【0128】

プローブは、図に示されている核酸分子の全長における、何れの配列とも対応することができる。したがって、それは5'非コード領域、コード領域、及び3'非コード領域から誘導することができる。しかしながら、すでに述べたように、フラグメントは本発明以前に公開されたフラグメントを含まないものと見なされる。

【0129】

核酸分子はまた、核酸分子の何れかの領域を増幅するPCRのプライマーとしても有用であり、必要な長さ及び配列のアンチセンス分子の合成においても有用である。

10

【0130】

核酸分子はまた、組み換えベクターの製造にも有用である。このようなベクターとしては、ペプチド配列の一部、又は全部を発現する発現ベクターが含まれる。ベクターはまた、挿入ベクターも含み、これは他の核酸分子中に統合され、例えば、細胞ゲノム中で遺伝子及び／又は遺伝子生成物の*in situ*発現を変化させるために用いられる。例えば、内生コード配列では、1つ以上の特異的に導入された変異を含むコード領域の一部、又は全部との相同的な組み換えを経て置換され得る。

【0131】

核酸分子はまた、タンパク質の抗原部分を発現することにも有用である。

20

【0132】

核酸分子はまた、*in situ*ハイブリダイゼーション法により、核酸分子の染色体の位置を決定するためのプローブとしても有用である。本発明の新規な薬剤一代謝タンパク質をコード化している遺伝子は、(図3に示されるように)ヒト染色体1上にマップされるゲノム成分上に位置しており、これはSTS及びBACマップデータのような多系統の証拠によって支持されている。

【0133】

核酸分子はまた、本発明の核酸分子の遺伝子調節領域を含むベクターの製造にも有用である。

【0134】

核酸分子はまた、ここに記載される核酸分子から生成されるmRNAの一部又は全部と対応しているリボザイムの設計にも有用である。

30

【0135】

核酸分子はまた、ペプチドの一部又は全部を発現しているベクターの製造にも有用である。

【0136】

核酸分子はまた、核酸分子及びペプチドの一部又は全部を発現している宿主細胞の製造にも有用である。

【0137】

核酸分子はまた、核酸分子及びペプチドの一部又は全部を発現している遺伝子組み換え動物の製造にも有用である。

40

【0138】

核酸分子はまた、核酸発現の存在、レベル、形態、分布を決定するためのハイブリダイゼーションプローブとしても有用である。図1の実験データによると、本発明の薬剤一代謝酵素タンパク質は、ヒトの胃、脳(乳児を含む)、子宮内膜腫瘍、前立腺、腎臓、副腎腫瘍、頭部／頸部、交感神経幹、胸部、及び肝細胞腫瘍で発現することが、仮想ノーザンブロット分析により示されている。PCRベースの組織スクリーニングパネルでは、脳で発現することも示されている。したがって、プローブは、細胞、組織、生体中の核酸分子の存在の検出、又はレベルの決定に用いることができる。レベルを決定される核酸はDNA又はRNAであり得る。したがって、ここに記載されるペプチドに対応するプローブは

50

、与えられた細胞、組織、生体中での発現及び／又は遺伝子コピー数の評価に用いることができる。これらの使用は、正常の場合と比較して、薬剤一代謝酵素タンパク質が増加又は減少していることに関係する障害の診断に適している。

【0139】

mRNAの*in vitro*検出技術には、ノーザンハイブリダイゼーション、及び*in situ*ハイブリダイゼーションが含まれる。DNAの*in vitro*検出技術には、サザンハイブリダイゼーション、及び*in situ*ハイブリダイゼーションが含まれる。

【0140】

プローブは、薬剤一代謝酵素タンパク質を発現する細胞又は組織の判定を行う診断テストキットの一部として用いることができる。これは、例えば、被験者の細胞サンプル中の、薬剤一代謝酵素をコード化した核酸分子、例えば、mRNA、又はゲノムDNAのレベルを測定するか、又は薬剤一代謝酵素遺伝子に変異していないか判定するものである。図1の実験データによると、本発明の薬剤一代謝酵素タンパク質は、ヒトの胃、脳（乳児を含む）、子宮内膜腫瘍、前立腺、腎臓、副腎腫瘍、頭部／頸部、交感神経幹、胸部、及び肝細胞腫瘍で発現することが、仮想ノーザンブロット分析により示されている。PCRベースの組織スクリーニングパネルでは、脳で発現することも示されている。

10

【0141】

核酸発現アッセイは、薬剤一代謝酵素核酸発現を変調する化合物を同定するための薬剤スクリーニングにおいて有用である。

20

【0142】

したがって、本発明は、薬剤一代謝酵素遺伝子の核酸発現、特に、細胞及び組織において薬剤一代謝酵素を媒介する生物学的、病理学的プロセスに関連した障害の治療に用いる化合物を同定する方法を提供するものである。図1の実験データによると、ヒトの胃、脳（乳児を含む）、子宮内膜腫瘍、前立腺、腎臓、副腎腫瘍、頭部／頸部、交感神経幹、胸部、及び肝細胞腫瘍で発現することが示されている。この方法には、典型的には、薬剤一代謝酵素核酸の発現を変調する化合物の能力についてアッセイを行うこと、及びその後、不必要な薬剤一代謝酵素核酸発現により特徴づけられる障害を治療するのに用いることができる化合物を同定することが含まれる。このアッセイは、細胞系及び無細胞系において行われることができる。細胞系のアッセイには、天然に薬剤一代謝酵素核酸を発現している細胞、又は特定の核酸配列を表現するために設計された組み換え細胞が含まれる。

30

【0143】

したがって、薬剤一代謝酵素遺伝子発現のモジュレータは、細胞と候補化合物とを接触させ、mRNAの発現を判定する方法により同定することができる。候補化合物の存在下での薬剤一代謝酵素mRNAの発現レベルは、候補化合物非存在下での薬剤一代謝酵素mRNAの発現レベルと比較される。この比較に基づいて、候補化合物は核酸発現のモジュレータとして同定され、例えば、異常核酸発現により特徴付けられる障害の治療に用いることができる。候補化合物存在下でのmRNAの発現が、非存在下のものと比較して統計的に著しく大きい場合、候補化合物は核酸発現の促進剤として同定される。候補化合物存在下でのmRNAの発現が、非存在下のものと比較して統計的に著しく小さい場合、候補化合物は核酸発現の阻害剤として同定される。

40

【0144】

本発明はさらに、薬剤一代謝酵素を発現する細胞及び組織における薬剤一代謝酵素核酸発現を変調する遺伝子モジュレータとしての薬剤スクリーニングを経て同定された化合物をターゲットとして用いる治療方法を提供するものである。図1の実験データによると、本発明の薬剤一代謝酵素タンパク質は、ヒトの胃、脳（乳児を含む）、子宮内膜腫瘍、前立腺、腎臓、副腎腫瘍、頭部／頸部、交感神経幹、胸部、及び肝細胞腫瘍で発現することが、仮想ノーザンブロット分析により示されている。PCRベースの組織スクリーニングパネルでは、脳で発現することも示されている。変調には、上方調整（例えば、活性化又はアゴニゼーション）、下方調整（抑制又はアンタゴニゼーション）の両者、又は核酸発現

50

が含まれる。

【0145】

あるいは、薬剤又は小分子がタンパク質を発現している細胞又は組織中で薬剤一代謝酵素発現を阻害するものである限りは、薬剤一代謝酵素核酸発現のモジュレータは、ここに記載されるスクリーニングアッセイを用いて同定される薬剤又は小分子であり得る。図1の実験データによると、ヒトの胃、脳（乳児を含む）、子宮内膜腫瘍、前立腺、腎臓、副腎腫瘍、頭部／頸部、交感神経幹、胸部、及び肝細胞腫瘍で発現することが示されている。

【0146】

核酸分子はまた、臨床試験又は治療方法において、薬剤一代謝酵素遺伝子の発現及び活性に対する変調化合物の効果をモニターするのに有用である。したがって、遺伝子発現パターンは、化合物、特に患者の耐性を向上させる化合物を用いた治療における、継続的効果のバロメータとなり得る。遺伝子発現パターンはまた、化合物に対する細胞の生理的反応を示すマーカーとなり得る。したがって、このようなモニタリングにより、化合物の投与量の増加、又は患者が耐性を示さない代替化合物の投与を行うことができる。同様に、核酸発現のレベルが望ましいレベルまで低下したならば、化合物の投与をこれに比例して減少することができる。

【0147】

核酸分子はまた、薬剤一代謝酵素核酸発現の質的变化、特に疾患に至る質的变化の診断アッセイにも有用である。核酸分子は、mRNAのような薬剤一代謝酵素遺伝子、及び遺伝子発現生成物における突然変異の検出に用いることができる。核酸分子は、薬剤一代謝酵素遺伝子において自然発生した遺伝子突然変異を検出し、その変異を持つ被験者が変異により生じる障害の危険性を有しているかどうかを判定するためのハイブリダイゼーションプローブとして用いることができる。突然変異には、欠失、付加、又は遺伝子中の1以上のヌクレオチドの置換、倒置又は転移のような染色体の組み換え、異常メチル化パターンのようなゲノムDNAの修飾、又は増幅のような遺伝子コピー数の変化が含まれる。機能障害に関連する薬剤一代謝酵素遺伝子の変異体の検出は、疾患が薬剤一代謝酵素タンパク質の過剰発現、過小発現、変異発現の結果生じる場合に、活性又は感受性の診断ツールを提供するものである。

【0148】

薬剤一代謝酵素遺伝子における突然変異をもたらしている個体は、種々の技術によって核酸レベルにおいて検出されることができる。図3には、本発明の薬剤一代謝酵素タンパク質をコード化している遺伝子において見出されたSNP情報を示す。挿入／欠失変異体（“indels”）を含むSNPは、45の異なるヌクレオチド位置で確認された。これらのSNPにより起こるアミノ酸配列の変化は、ユニバーサル遺伝子コード、及び図2に示されるタンパク質配列を用いて容易に確認することができる。エキソン、イントロン、及びORFの外側のそれぞれのSNPの位置は、それぞれのSNPより得られるDNA位置、及びその性質より得られる開始／終止、エキソン及びイントロン座標を用いて容易に決定することができる。本発明の新規な薬剤一代謝タンパク質をコード化している遺伝子は、（図3に示されるように）ヒト染色体1上にマップされるゲノム成分上に位置しており、これはSTS及びBACマップデータのような多系統の証拠によって支持されている。ゲノムDNAは、直接又は予めPCRを用いて増幅した後で分析されることができる。RNA又はcDNAも、同様にして用いることができる。ある使用においては、突然変異の検出は、例えば、アンカーPCR、RACEPCRのような、ポリメラーゼ連鎖反応（PCR）（例えば、U. S. Patent No. 4, 683, 195, 及び 4, 683, 202 参照）、又は他のものとして、リゲーション連鎖反応（LCR）（例えば、Landegran et al., Science 241: 1077-1080 (1988); 及び Nakazawa et al., PNAS 91: 360-364 (1994) 参照）において、プローブ／プライマーの使用に関連し、特に後者は遺伝子中の変異位置の同定に有用である（Abravaya et al., Nucleic Acids Res. 23: 675-682 (1995) 参照）。この方法には、患者から

10

20

30

40

50

細胞サンプルを収集する工程と、サンプルの細胞から核酸（例えば、ゲノム、mRNA又はその両方）を単離する工程と、遺伝子（もし存在すれば）のハイブリダイズ及び増幅が起こりうる条件下で遺伝子に特異的にハイブリダイズする1以上のプライマーと核酸とを接触させる工程と、増幅生成物の存在又は非存在を検出するか、又は増幅生成物のサイズを検出し、コントロールサンプルの長さと比較する工程を含む。欠失及び挿入は、増幅生成物のサイズの変化を、正常な遺伝子型と比較することにより検出することができる。点突然変異は、増幅DNAと正常なRNA、又はアンチセンスのDNA配列とハイブリダイズすることによって確認することができる。

【0149】

あるいは、薬剤一代謝酵素遺伝子の突然変異は、例えば、ゲル電気泳動により決定される制限酵素消化パターンの変更により、直接的に確認することができる。

10

【0150】

さらに、配列特定リボザイム（U. S. Patent No. 5, 498, 531）は、リボザイム開裂部位の成長又は減少により、特定の変異の存在の評点のために用いることができる。完全に一致する配列は、ヌクレアーゼ開裂消化分析評価、又は融点の違いによって、不一致の配列から分離することができる。

【0151】

特定位置での配列変化は、RNase及びS1保護、又は化学開裂法のようなヌクレアーゼ保護アッセイによって評価することができる。さらに、変異体薬剤一代謝酵素遺伝子と野生型遺伝子との配列の相違は、直接DNA配列解析によって決定することができる。種々の自動化された配列解析手段は、診断アッセイ（Naevé, C. W., (1995) *Biotechniques* 19: 448）の実行において有用であり、これらには、マスペクトルによる配列解析（例えば、PCT International Publication No. WO 94/16101; Cohen et al., *Adv. Chromatogr.* 36: 127-162 (1996), and Griffin et al., *Appl. Biochem. Biotechnol.* 38: 147-159 (1993) 参照）も含まれる。

20

【0152】

遺伝子中の突然変異を検出する他の技術の例としては、RNA/RNA、又はRNA/DNA二重鎖から、不一致の塩基を検出するために、開裂試薬から保護する方法（Myers et al., *Science* 230: 1242 (1985), Cotton et al., *PNAS* 85: 4397 (1988), Saleeba et al., *Meth. Enzymol.* 217: 286-295 (1992))、変異体と野生型の核酸の電気泳動移動度を比較する方法（Orita et al., *PNAS* 86: 2766 (1989); Cotton et al., *Mutat. Res.* 285: 125-144 (1993); and Hayashi et al., *Genet. Anal. Tech. Appl.* 9: 73-79 (1992))、及び変性剤の勾配を含んだポリアクリルアミドゲル中での変異体又は野生型のフラグメントの動きを、勾配ゲル電気泳動を用いてアッセイする方法（Myers et al., *Nature* 313: 495 (1985))が含まれる。点突然変異を検出する他の技術の例としては、選択的オリゴヌクレオチドハイブリダイゼーション、選択的増幅、及び選択的プライマー伸長が含まれる。

30

40

【0153】

核酸分子は、治療方法としての効果を持つにも関わらず、必ずしも疾患を引き起こすわけではない遺伝子型のための、個人テストにおいても有用である。このため、核酸分子は、個人の遺伝子型と、治療に用いられる化合物に対する個人の応答との相関（薬理ゲノム相関）についての研究に用いることができる。したがって、ここに記載されている核酸分子は、治療のための適切な化合物及び投与量を選択するために、個人の薬剤一代謝酵素遺伝子の変異についての評価に用いることができる。図3には、本発明の薬剤一代謝酵素タンパク質をコード化している遺伝子において見出されたSNP情報を示す。挿入/欠失変異

50

体 (“ i n d e l s ”) を含む S N P は、45 の異なるヌクレオチド位置で確認された。これらの S N P により起こるアミノ酸配列の変化は、ユニバーサル遺伝子コード、及び図 2 に示されるタンパク質配列を用いて容易に確認することができる。エキソン、イントロン、及び O R F の外側のそれぞれの S N P の位置は、それぞれの S N P より得られる D N A 位置、及びその性質より得られる開始／終止、エキソン及びイントロン座標を用いて容易に決定することができる。

【0154】

このように、治療に影響する遺伝子変異を示す核酸分子は、個人におけるテイラー治療に用いることのできる診断ターゲットを提供するものである。したがって、これらの多形性を含んだ組み換え細胞及び組み換え動物の製造は、治療化合物及び薬剤投与についての効果的な臨床設計を可能とする。

10

【0155】

核酸分子は、細胞、組織、及び生体における薬剤一代謝酵素遺伝子発現を制御するためのアンチセンス構成物として有用である。D N A アンチセンスの核酸分子は、転写に関連する遺伝子の部位に対して相補的になるよう設計され、それ故に、薬剤一代謝酵素タンパク質の生成と転写が防がれる。アンチセンスの R N A 又は D N A 核酸分子は m R N A へとハイブリダイズされ、これにより薬剤一代謝酵素タンパク質中での m R N A の翻訳がブロックされる。

【0156】

あるいは、ある種のアンチセンス分子は、薬剤一代謝酵素核酸の発現を減少させる不活性化 m R N A に用いることができる。したがって、これらの分子は、異常、又は不必要な薬剤一代謝酵素核酸の発現により特徴づけられる障害の治療に用いることができる。この技術は、翻訳される m R N A の能力が減少した m R N A の 1 以上の領域に相補的なヌクレオチド配列を含んだリボザイムによる開裂に関連している。可能な領域としては、コード領域、特に、基質結合のような、薬剤一代謝酵素タンパク質の触媒活性及び他の機能活性に対応したコード領域が含まれる。

20

【0157】

核酸分子はまた、薬剤一代謝酵素遺伝子発現において異常な細胞を持つ患者の遺伝子治療のためのベクターを提供するものである。このため、組み換え細胞には、e x v i v o で調製され患者に戻される細胞が含まれ、個人の体内に導入されてそこで個人の治療のために必要とされる薬剤一代謝酵素タンパク質を生成する。

30

【0158】

本発明は、生物学的サンプル中の薬剤一代謝酵素核酸の存在を検出するために抗体を用いたキットも包含する。図 1 の実験データによると、本発明の薬剤一代謝酵素タンパク質は、ヒトの胃、脳（乳児を含む）、子宮内膜腫瘍、前立腺、腎臓、副腎腫瘍、頭部／頸部、交感神経幹、胸部、及び肝細胞腫瘍で発現することが、仮想ノーザンブロット分析により示されている。P C R ベースの組織スクリーニングパネルでは、脳で発現することも示されている。例えば、キットは、ラベル化された、又はラベル化可能な核酸、又は生物学的サンプル中で薬剤一代謝酵素核酸を検出することのできる試薬を含み、；サンプル中の薬剤一代謝酵素核酸量を決定する手段；標準サンプルの薬剤一代謝酵素核酸量と比較する手段、とを含むことができる。この化合物又は試薬は適当な容器に封入することができる。このキットは、さらに薬剤一代謝酵素タンパク質 m R N A 又は D N A の検出キットとして使用するための説明を含むことができる。

40

【0159】

核酸アレイ

本発明はさらに、核酸検出キットを提供するものであり、これらは、例えば、図 1 及び 3 (S E Q I D N O . 1 及び 3) に示される配列情報に基づいた核酸分子のアレイ又はマイクロアレイである。

【0160】

ここで用いられている、「アレイ」又は「マイクロアレイ」は、紙、ナイロン又は他の膜

50

、フィルタ、チップ、ガラススライド、又は他の適当な固形支持体のような基盤の上に合成された、別個のポリヌクレオチド、又はオリゴヌクレオチドのアレイのことである。1つの例として、マイクロアレイは、US Patent 5, 837, 832, Chee et al., PCT application W095/11995 (Chee et al.), Lockhart, D. J. et al. (1996; Nat. Biotech. 14: 1675-1680) 及び Schena, M. et al. (1996; Proc. Natl. Acad. Sci. 93: 10614-10619) に記載される方法にしたがって調製、使用され、これらの全ては参考としてここに折り込まれる。あるいは、このようなアレイは、Brown et al., US Patent No. 5, 807, 522. に記載される方法により製造される。

10

【0161】

マイクロアレイ又は検出キットは、好適には、多数の特異的な単鎖の核酸配列を構成し、通常は合成アンチセンスのオリゴヌクレオチドか、又はcDNAのフラグメントのどちらかが固体支持体上に固定される。オリゴヌクレオチドは、好適には約6～60のヌクレオチド長、より好適には15～30のヌクレオチド長、最も好適には20～25のヌクレオチド長である。あるタイプのマイクロアレイ又は検出キットには、7～20のヌクレオチド長のオリゴヌクレオチドのみを使うことが好適であり得る。マイクロアレイ又は検出キットは、既知の5'又は3'配列を含んだオリゴヌクレオチド、全長配列を含んだオリゴヌクレオチド、又は配列長さの特定領域から選択された特異なオリゴヌクレオチドを含むものであり得る。マイクロアレイ又は検出キットにおいて用いられるポリヌクレオチドは、遺伝子又は対象となる遺伝子において特異的なオリゴヌクレオチドであり得る。

20

【0162】

マイクロアレイ又は検出キットにおいて、既知の配列のオリゴヌクレオチドを製造するために、対象となる遺伝子（又は、本発明により確認されたORF）は、典型的には、核酸配列の5'から開始、又は3'で終了するコンピュータアルゴリズムを用いて試験される。典型的なアルゴリズムでは、遺伝子に特異的な長さに規定されたオリゴマーが同定され、ハイブリダイゼーションに好適な範囲にGC成分を持ち、ハイブリダイゼーションの妨害となると予測される二次構造を持たない。ある条件では、マイクロアレイ又は検出キットにおいて、オリゴヌクレオチドのペアを用いることが好適であり得る。オリゴヌクレオチドの「ペア」は、好適には配列の中央に位置している1つのヌクレオチドを除いては、同一である。ペアの2つ目のオリゴヌクレオチド（一方とは不一致）はコントロールとして用いられる。オリゴヌクレオチドペアの数は、2から100万の間である。オリゴマーは、光誘導化学プロセスを用いて、基盤上の指定領域で合成される。基盤は、紙、ナイロン又は他の膜、フィルタ、チップ、ガラススライド、又は他の適当な固形支持体である。

30

【0163】

他方、オリゴヌクレオチドは、PCT application W095/251116 (Baldeschweiler et al.) に記載されているように、化学カップリング手段、及びインクジェットアプリケーション装置を用いて基盤の表面上で合成され、これらの全ては参考としてここに折り込まれる。他の観点では、ドット（又はスロット）プロットに似た「格子」アレイでは、真空システム、加熱、UV、機械的又は化学的結合工程を用いて、cDNA、又はオリゴヌクレオチドを基板の表面上に配置、結合させることができる。上記のようなアレイは、手工、又は利用可能な装置（スロットプロット、又はドットプロット装置）、材料（適当な固形支持体）、及び機械（ロボット装置を含む）を用いて製造され、また、8、24、96、384、1536、6144又はこれ以上、又は商業的に用いられる装置に効果的に使用される2から100万の間の他の数のオリゴヌクレオチドを含んでも良い。

40

【0164】

マイクロアレイ又は検出キットを用いてサンプルの分析を行うために、生物サンプルから得られたRNA又はDNAは、ハイブリダイゼーションプローブ中に調製される。mRNAが単離され、そしてcDNAが調製され、アンチセンスのRNA（aRNA）を調製す

50

るためのテンプレートとして用いられる。aRNAは蛍光性ヌクレオチドの存在化で増幅し、ラベル化されたプローブがマイクロアレイ又は検出キットにおいてインキュベートされ、そして、プローブの配列がマイクロアレイ又は検出キット中の相補的なオリゴヌクレオチドとハイブリダイズされる。インキュベート条件は、正確に相補的に一致しているか、又は各種程度の相補性でハイブリダイゼーションが起こるように調節される。ハイブリダイズしていないプローブを除去した後、蛍光のレベルとパターンを判定するためにスキャナが用いられる。スキャンされたイメージは、マイクロアレイ又は検出キット上の、相補性の程度、及び各々のオリゴヌクレオチド配列の相対的な量を決定するために試験される。生物学的サンプルは、体液（例えば血、尿、唾液、痰、胃液、その他）、培養細胞、生体組織検査、又は他の組織調製品から得られる。検出システムでは、全ての異なる配列において、ハイブリダイゼーションの存在、非存在、及び量を、同時に計測するのに用いられる。このデータは、サンプル中での、配列、発現パターン、変異、変異体、又は多形性といった、大規模な相関性の研究に用いられる。

10

【0165】

本発明は、このようなアレイを用いて、本発明の薬剤－代謝酵素タンパク質／ペプチドの発現を同定するための方法を提供するものである。詳細には、このような方法は、テストサンプルと一つ以上の核酸分子とのインキュベートと、テストサンプル中の成分と核酸分子との結合についてのアッセイとが含まれる。このようなアッセイは、少なくとも遺伝子の一つが本発明の遺伝子及び／又は本発明の薬剤－代謝酵素遺伝子の対立変異体である、多くの遺伝子を含むアレイに関連している。図3には、本発明の薬剤－代謝酵素タンパク質をコード化している遺伝子において見出されたSNP情報を示す。挿入／欠失変異体（“indels”）を含むSNPは、45の異なるヌクレオチド位置で確認された。これらのSNPにより起こるアミノ酸配列の変化は、ユニバーサル遺伝子コード、及び図2に示されるタンパク質配列を用いて容易に確認することができる。エキソン、イントロン、及びORFの外側のそれぞれのSNPの位置は、それぞれのSNPより得られるDNA位置、及びその性質より得られる開始／終止、エキソン及びイントロン座標を用いて容易に決定することができる。

20

【0166】

テストサンプルと核酸分子のインキュベートの条件は様々である。インキュベーション条件は、使用されるアッセイの形式、使用される検出方法、及びアッセイに用いられる核酸分子のタイプ及び性質に依存する。一般的に入手可能なハイブリダイゼーション、増幅、又はアレイアッセイの形式の何れかを認識している当業者は、ここに記載されるヒトゲノムの新規フラグメントを用いるに当って、容易に適用を行うことができる。このようなアッセイの例は、Chard, T., An Introduction to Radioimmunoassay and Related Techniques, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G. R. et al., Techniques in Immunocytochemistry, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1985)に記載されている。

30

40

【0167】

本発明のテストサンプルには、細胞、タンパク質、細胞からの膜抽出物が含まれる。上記の方法に用いられるテストサンプルは、アッセイの形式、検出方法の性質、及びアッセイのサンプルとして用いられる組織、細胞、又はその抽出物に基づいて変化する。核酸抽出物又は細胞抽出物の調製は、当業者において周知であり、用いられるシステムと調和するサンプルを得ることにより、容易に適用することができる。

50

【0168】

本発明の他の例としては、本発明のアッセイを行うために必要な試薬を含むキットが提供される。

【0169】

特に本発明は、(a)ここに記載されるヒトゲノムのフラグメントと結合することのできる1以上の核酸分子を含む第一の容器、(b)1以上の洗浄試薬、結合核酸を検出することのできる試薬を含む他の1以上の容器、とを含む1以上の容器に区分され、封入されたキットを提供するものである。

【0170】

詳細には、区分されたキットとしては、試薬が別々の容器に含まれているキットが含まれる。10
このような容器としては、小さいガラスの容器、プラスチック容器、帯状のプラスチック、ガラス又は紙、又はシリカのようなアレイ材料が含まれる。このような容器は、サンプルと試薬が混合して汚染しないように、1つの区分から他の区分へと試薬を効率的に移動することができ、またそれぞれの容器の試薬又は溶液は他の容器へと定量的に添加することができる。このような容器には、テストサンプルを入れる容器、核酸プローブを含む容器、洗浄試薬（例えば、リン酸塩緩衝液、T r i s -緩衝液等）を含む容器、及び結合プローブを検出に用いられる試薬を含む容器を含む。当業者は、本発明にかかる従来未知の薬剤-代謝酵素遺伝子を認識し、ここに開示されている配列情報を用いて定型的に確認することができ、さらにこれを当業者において周知の確立されたキット形態、特に発現アレイに組み込むことができる。 20

【0171】

ベクター／宿主細胞

本発明はまた、ここに記載される核酸分子を含んだベクターを提供するものである。「ベクター」という用語は、ビヒクルのことを指し、好適には核酸分子であり、核酸分子の輸送をすることができるもののことである。ベクターが核酸分子である場合、核酸分子はベクターの核酸と共有結合している。本発明のこの観点では、ベクターには、プラスミド、単鎖又は二重鎖のファージ、単鎖又は二重鎖のRNA又はDNAのウイルス性ベクター、又はBAC、PAC、YAC、ORMACのような人工染色体が含まれる。

【0172】

ベクターは宿主細胞中に染色体外の成分として保持され、そこで核酸分子の付加的なコピーを複製及び生成する。あるいは、ベクターは宿主細胞のゲノム中に組み込まれ、宿主細胞の複製の際に核酸分子の付加的なコピーを生成する。 30

【0173】

本発明は、核酸分子の保持のためのベクター（クローニングベクター）、又は核酸分子の発現のためのベクター（発現ベクター）を提供するものである。このベクターは、原核生物細胞又は真核生物細胞、又はその両方で機能することができる（シャトルベクター）。

【0174】

発現ベクターは、ベクター中で核酸分子と有効に結合したcis作用性調節領域を含み、これにより宿主細胞中での核酸分子の転写が可能となる。この核酸分子は、転写に影響を及ぼす核酸分子と分離されて、宿主細胞に導入されることができる。したがって、第二の核酸分子は、ベクターからの核酸分子の転写を行うcis調節制御領域と相互作用するトランス作用性因子を提供し得る。あるいは、トランス作用性因子は宿主細胞により提供され得る。最終的に、トランス作用性因子は、ベクター自身から作り出すことができる。しかしながら、いくつかの例では、核酸分子の転写及び／又は翻訳は無細胞系でも起こり得る。 40

【0175】

ここに記載される核酸分子の調整配列は、目的のmRNA転写のためのプロモーターを含んで有効に結合されることができる。これらには、バクテリオファージλからの左部プロモーター、E. coliからのlac、TRP及びTACプロモーター、SV40からの初期及び後期のプロモーター、CMVの極初期のプロモーター、アデノウイルスの初期及 50

び後期のプロモーター、及びレトロウイルスのLTRが含まれるが、これに限定されるものではない。

【0176】

転写を促進する制御領域に加えて、発現ベクターはまた、リプレッサー結合部位やエンハンサーのような転写を調整する領域を含むものであり得る。この例としては、SV40エンハンサー、サイトメガロウイルスの極初期のエンハンサー、ポリオーマエンハンサー、アデノウイルスエンハンサー、レトロウイルスLTRエンハンサーが含まれる。

【0177】

転写の開始及び制御領域を含む場合に加えて、発現ベクターはまた、転写のためのリボソーム結合部位である転写領域における、転写終了のために必要な配列を含むことができる。他の発現調整制御成分としては、ポリアデニル化信号と同様に、開始及び終止コドンが含まれる。当業者は、発現ベクターに有用な多数の調整配列を知り得る。このような調整配列は、例えば、Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989)に記載されている。

【0178】

各種の発現ベクターは、核酸分子の発現に用いることができる。このようなベクターには、染色体、エピソーム、ウイルス由来のベクター、例えば、バクテリアプラスミド、バクテリオファージ、酵母エピソーム、人工酵母染色体のような酵母染色体成分、バクロウイルス、SV40のようなパポバウイルス、バクシニアウイルス、アデノウイルス、ポクスウイルス、シュードラビスウイルス、及びレトロウイルスのようなウイルス由来のベクターが含まれる。ベクターはまた、これらの起源の組み合わせから誘導することができ、例えば、コスミド及びファージミドのようなプラスミドとバクテリオファージの遺伝子成分から誘導することができる。原核及び真核生物の宿主細胞のための適切なクローニングベクター及び発現ベクターは、Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989)に記載されている。

【0179】

調整配列では、1以上の宿主細胞の構成的な発現（すなわち組織特異性）、又は、温度、養分添加、又はホルモンや他のリガンドのような外生の要因による1以上の細胞タイプでの指示的な発現を提供するものである。原核及び真核生物の宿主細胞において構成的、及び指示的に発現する種々のベクターは、当業者において周知である。

【0180】

核酸分子は、周知の方法によってベクター核酸内に導入されることができる。通常、最終的に発現するDNA配列は、DNA配列と発現ベクターが1以上の限定酵素により開裂し、その後フラグメントが共に結合することによって、発現ベクターと結合される。制限酵素の消化及び結合の手順は、当業者において周知である。

【0181】

適切な核酸分子を含んでいるベクターは、公知の技術を用いて、増殖又は発現のために適切な宿主細胞内へ導入することができる。バクテリア細胞には、E. coli、Streptomyces、及びSalmonella typhimuriumが含まれるがこれに限定されるものではない。真核生物細胞には、酵母、Drosophilaのような昆虫細胞、COS及びCHO細胞のような動物細胞、及び植物細胞が含まれるが、これらに限定されるものではない。

【0182】

ここに記載されているように、融合タンパク質としてのペプチドの発現が望ましい。したがって、本発明はペプチドの生成が可能な融合ベクターを提供するものである。融合ベクターは組み換えタンパク質の発現及び溶解性を向上することができ、また、例えば、アフ

10

20

30

40

50

ィニティー精製のためのリガンドの作用によって、タンパク質精製を向上することができる。タンパク質分解性開裂部位は融合部分との結合位置に導入され、このために、目的となるペプチドは最終的に融合部分から分離される。タンパク質分解酵素としては、ファクターXa、スロンビン、エンテロキナーゼが含まれるが、これに限定されるものではない。典型的な融合発現ベクターとしては、グルタチオンS-転移酵素(GST)、マルトースE結合タンパク質、又はタンパク質Aのそれぞれをターゲット組み換えタンパク質に融合した、pGEX (Smith et al., Gene 67:31-40 (1988))、pMAL (New England Biolabs, Beverly, MA)、pRIT5 (Pharmacia, Piscataway, NJ)が含まれるが、これに限定されるものではない。好適な指示的非融合E. coli発現ベクターの例としては、

10

【0183】

組み換えタンパク質の発現は、遺伝子背景を与えることによって、宿主バクテリアにおいて最大化することができ、宿主細胞は組み換えタンパク質のタンパク質分解性の開裂欠損能力を有する (Gottesman, S., Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, California (1990) 119-128)。あるいは、対象となる核酸分子の配列は、例えば、E. coliのような特定の宿主細胞のために優先的に使用されるコドンとなるように変更されることができる (Wada et al., Nucleic Acids Res. 20:2111-2118 (1992))。

20

【0184】

核酸分子はまた、酵母において作用する発現ベクターにより発現されることもできる。S. cerevisiaeのような酵母中で発現するベクターの例としては、pYepSec1 (Baladari, et al., EMBO J. 6:229-234 (1987))、pMFa (Kurjan et al., Cell 30:933-943 (1982))、pJRY88 (Schultz et al., Gene 54:113-123 (1987))、pYES2 (Invitrogen Corporation, San Diego, CA)が含まれる。

30

【0185】

核酸分子はまた、例えば、バクロウイルス発現ベクターを用いて、昆虫細胞内で発現されることもできる。培養昆虫細胞 (例えば、Sf9細胞) 中のタンパク質の発現に利用されるベクターには、pAcシリーズ (Smith et al., Mol. Cell Biol. 3:2156-2165 (1983))、及びpVLシリーズ (Lucklow et al., Virology 170:31-39 (1989))が含まれる。

【0186】

本発明のある例においては、ここに記載されている核酸分子は、哺乳類発現ベクターを用いて哺乳類の細胞内で発現される。哺乳類発現ベクターの例としては、pCDM8 (Seed, B. Nature 329:840 (1987))、及びpMT2PC (Kaufman et al., EMBO J. 6:187-195 (1987))が含まれる。

40

【0187】

ここに列記されている発現ベクターとしては、核酸分子を発現するために有用であり、当業者が利用可能なベクターとして周知のもののみが示されている。ここに記載されている核酸分子の維持増殖、又は発現に好適な他のベクターは、当業者において周知である。これらは、例えば、Sambrook, J., Fritsh, E. F., and Maniatis, T. Molecular Cloning: A Laboratory

50

Manual, 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989に記載されている。

【0188】

ここに記載されている核酸配列がベクター中に逆方向にクローンされたベクターは、アンチセンスRNAの転写を許す調整配列に結合可能であるが、本発明は、また、このようなベクターも包含するものである。このように、アンチセンス転写は、ここに記載され、コード、非コード領域の両方が含まれている核酸分子配列の、全部又は一部を生成することができる。そして、このアンチセンスのRNAの発現は、センスRNAの発現（調整配列、構成的、又は指示的発現、組織特異発現）に関して、前記した各パラメータに対応する。

10

【0189】

本発明はまた、ここに記載されるベクターを含む組み換え宿主細胞に関連するものである。したがって、宿主細胞には、原核生物細胞、酵母のような低真核生物細胞、昆虫細胞のような他の真核生物細胞、及び哺乳類の細胞のような高真核生物細胞が含まれる。

【0190】

組み換え宿主細胞は、当業者が容易に利用可能な技術により、ここに記載されるように構成されるベクターを細胞中に導入することにより、調製することができる。これらには、リン酸カルシウムトランスフェクション、DEAE-デキストラン媒介トランスフェクション、カチオン性脂質媒介トランスフェクション、エレクトロポレーション、トランスダクション、インフェクション、リポフェクション、及びSambrook, et al.

20

(Molecular Cloning: A Laboratory Manual, 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989)に記載されているような他の技術が含まれるが、これらに限定されるものではない。

【0191】

宿主細胞は、1以上のベクターを含むことができる。このため、異なるヌクレオチド配列が、同じ細胞の異なるベクター中に導入されることができる。同様に、核酸分子は、単独で、又は発現ベクターのトランス作用因子を与えているような、関連のない他の核酸分子と共に導入されることができる。1以上のベクターが細胞内に導入される場合、ベクターは単独で導入されるか、共に導入されるか、又は核酸分子ベクターに結合されることができる。

30

【0192】

バクテリオファージ及びウィルスベクターの場合、これらは標準的なインフェクション及びトランスダクションの操作により、封入又はカプセル化されたウイルスとして細胞内に導入されることができる。ウィルスベクターは、複製可能、又は複製欠陥であり得る。ウイルスの複製に欠陥がある場合、複製は欠陥を補完する機能が提供される宿主細胞内で起こり得る。

【0193】

ベクターは一般に、組み換えベクターの構成物を含む細胞の部分母集団の選択を可能とする選択性マーカを含む。このマーカは、ここに記載される核酸分子を含む同一のベクター内か、又は別のベクター中に含まれることができる。マーカには、原核生物宿主細胞のためのテトラサイクリン又はアンピシリン抵抗遺伝子、及び真核生物宿主細胞のためのジヒドロフォレート還元酵素又はネオマイシン耐性が含まれる。しかしながら、表現型特性の選択性を提供するマーカは何れの場合にも有効である。

40

【0194】

成熟タンパク質は、バクテリア、酵母、哺乳類の細胞、及び他の細胞において、適切な調整配列の制御下で生成されることができるが、無細胞系転写及び翻訳システムもまた、ここに記載されるDNA構成物から誘導されるRNAを用い、これらのタンパク質を生成す

50

るために用いることができる。

【0195】

ペプチドの分泌が必要とされる場合、これが薬剤一代謝酵素のようなタンパク質を含むマルチ膜貫通ドメイン内で達成されることは難しく、適切な分泌信号がベクター中に組み込まれる。信号配列は、これらのペプチドに内生であるか、又はペプチドに非相同であり得る。

【0196】

ペプチドが媒体内で分泌されない場合、典型的には薬剤一代謝酵素の場合、タンパク質は、凍結融解、超音波処理、機械的破壊、分解試薬等の標準的な破壊操作によって、宿主細胞から単離されることができる。ペプチドは、硫酸アンモニウム沈降、酸抽出、又はアニオン又はカチオン交換クロマトグラフィ、ホスホセルロースクロマトグラフィ、疎水性相互作用クロマトグラフィ、アフィニティクロマトグラフィ、ヒドロキシルアパタイトクロマトグラフィ、レクチンクロマトグラフィ、又は高速液体クロマトグラフィを含む、公知の精製方法によって、回収、精製されることができる。

10

【0197】

また、ここに記載されているペプチドの組み換え製造における宿主細胞に依存して、ペプチドは種々のグリコシル化パターンを持ち、細胞に依存してバクテリア内で製造される際にグリコシル化されないかもしれないことが理解される。さらにペプチドは、ホストを媒介する過程の結果として、いくつかの場合で最初に修飾されたメチオニンを含むものであり得る。

20

【0198】

ベクター及び宿主細胞の使用

ここに記載されているペプチドを発現している組み換え宿主細胞には、種々の使用用途がある。まず、この細胞は薬剤一代謝酵素タンパク質、又はペプチドの生成に有用であり、薬剤一代謝酵素タンパク質又はフラグメントを必要量生成するために、さらに精製を行うことができる。このため、発現ベクターを含む宿主細胞は、ペプチドの生成に有用である。

【0199】

宿主細胞は、薬剤一代謝酵素タンパク質、又は薬剤一代謝酵素タンパク質フラグメントに関連している細胞系のアッセイ、例えば上記したもの、同様に当業者において周知の他の形態のものの実行において有用である。このため、天然の薬剤一代謝酵素タンパク質を発現している組み換え宿主細胞は、薬剤一代謝酵素タンパク質機能を促進又は阻害する化合物のアッセイに有用である。

30

【0200】

宿主細胞はまた、機能的な影響を受ける薬剤一代謝酵素タンパク質変異体を同定するために有用である。変異が自然に生じて病理を引き起こすような場合、突然変異を含む宿主細胞は、天然の薬剤一代謝酵素タンパク質の効果を示さずに、薬剤一代謝酵素タンパク質変異体に要求される効果（例えば、機能を促進、又は阻害）を持つ化合物のアッセイに有用である。

【0201】

遺伝子的に工作された宿主細胞は、さらにヒト以外の遺伝子組み換え動物を生産するために用いることができる。遺伝子組み換え動物は、好適には哺乳類であり、例えば、1以上の細胞が組み換え遺伝子を含んだ、ラット又はマウスのような齧歯動物である。組み換え遺伝子は、成長中の遺伝子組み換え動物の細胞のゲノムに組み込まれ、1以上の細胞型又は組織において、成熟した動物のゲノム中に残存する外生のDNAである。これらの動物は、薬剤一代謝酵素タンパク質の機能の研究、及び薬剤一代謝酵素タンパク質活性のモジュレータの同定及び評価に有用である。遺伝子組み換え動物の他の例としては、ヒト以外の霊長類、羊、犬、牛、ヤギ、鶏、及び両生類が含まれる。

40

【0202】

遺伝子組み換え動物は、例えば、マイクロインジェクション、レトロウイルス感染によっ

50

て、受精卵母細胞の雄性前核細胞内に核酸分子を導入し、卵母細胞が偽妊娠性の雌性育成動物中で育成されることにより作製される。何れの薬剤一代謝酵素タンパク質ヌクレオチド配列も、マウスのようなヒト以外の動物のゲノム中に組み換え遺伝子として導入されることができる。

【0203】

発現ベクターに有用な調整配列、又は他の配列は、何れも組み換え遺伝子配列の一部を形成することができる。イントロン配列及びポリアデニル化信号が、すでに含まれていない場合には、これも含まれる。組織特異性調整配列は、特定の細胞に対し薬剤一代謝酵素タンパク質が直接発現するために、組み換え遺伝子に有効に結合されることができる。

【0204】

受胎操作及びマイクロインジェクションを通して、遺伝子組み換え動物を生産する方法、特にマウスのような動物を用いる方法は、当業界において一般化されており、例えば、U. S. Patent Nos. 4, 736, 866、及び 4, 870, 009, by Leder et al., U. S. Patent No. 4, 873, 191 by Wagner et al. and in Hogan, B., Manipulating the Mouse Embryo, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y., 1986) に記載されている。また、同様の方法が、他の遺伝子組み換え動物の生産のために用いられている。最初の遺伝子組み換え動物は、ゲノム中の組み換え遺伝子の存在及び／又は動物の組織や細胞内での遺伝子組み換えmRNAの発現に基づいて確認されることができる。最初の遺伝子組み換え動物は、その後、さらに組み換え遺伝子を有する動物を繁殖するために用いられることができる。その上、組み換え遺伝子を有している遺伝子組み換え動物は、さらに他の組み換え遺伝子を有する他の遺伝子組み換え動物へと生育されることができる。遺伝子組み換え動物はまた、ここに記載されている相同的な組み換え宿主細胞を用いて製造された、全ての動物又は動物の組織が含まれる。

【0205】

他の例では、ヒト以外の遺伝子組み換え動物は、組み換え遺伝子の調節された発現を行う選択システムを含むものとして生産されることができる。このようなシステムの1つの例は、バクテリオファージP1のcre/loxPリコンビナーゼシステムである。cre/loxPリコンビナーゼシステムについての記載は、例えば、Lakso et al., PNAS 89: 6232-6236 (1992) 参照。リコンビナーゼシステムのもう一つの例は、S. cerevisiaeのFLPリコンビナーゼシステムである (O'Gorman et al., Science 251: 1351-1355 (1991))。cre/loxPリコンビナーゼシステムが組み換え遺伝子の発現の調節に用いられる場合は、動物において、Creリコンビナーゼ及び選択されたタンパク質の両方をコード化している組み換え遺伝子が含まれていることが必要である。このような動物は、例えば、一方は選択されたタンパク質をコード化した組み換え遺伝子を持ち、他方はリコンビナーゼをコード化した組み換え遺伝子を持った2つの組み換え遺伝子動物を交配させることにより、「二重」遺伝子組み換え動物を構成することによって提供される。

【0206】

ここに記載されるヒト以外の遺伝子組み換え動物のクローンは、また、Wilmut, I. et al., Nature 385: 810-813 (1997) 及び、PCT International Publication Nos. WO 97/07668 and WO 97/07669に記載されている方法に従って生産されることができる。簡単に述べると、遺伝子組み換え動物からの細胞、例えば体細胞は、単離されて、成長サイクルから出てG₀相に入れられるように誘導することができる。静止細胞は、例えば、電気パルスの使用によって、単離された静止細胞と同種の動物の細胞核を取り除かれた卵母細胞に融合されることができる。再構成された卵母細胞は、桑実胚又は芽細胞に発達するように培養され、その後、偽妊娠性の雌性育成動物中に移される。この雌性育成動物から誕生する子孫は、細胞、例えば体細胞を単離した動物のクローンとなる。

【0207】

ここに記載されているペプチドを発現する組み換え細胞を含んだ遺伝子組み換え動物は、*in vivo*の環境で、ここに記載したようなアッセイを行うために有用である。したがって、生体内に存在し、基質結合、薬剤一代謝酵素タンパク質活性化、信号伝達に影響を与えている各種の生理学的ファクターは、*in vitro*の無細胞系又は細胞系のアッセイでは明らかにならないかもしれない。このため、これらは、基質相互作用、薬剤一代謝酵素タンパク質機能及び基質相互作用に対する特定の変異体薬剤一代謝酵素タンパク質の影響、及びキメラ薬剤一代謝酵素タンパク質の影響を含む薬剤一代謝酵素タンパク質機能を、*in vivo*でアッセイするための、ヒト以外の遺伝子組み換え動物を提供するために有用である。また、実質的に又は完全に一つ以上の薬剤一代謝酵素タンパク質機能

10

【0208】

本明細書において、上に記載された全ての刊行物及び特許は、ここに参考として折り込まれている。本発明に記載された方法及びシステムの各種修正及び変形は、本発明の範囲及び精神から外れない限り、当業者において明らかなるものである。本発明は、特定の好適な具体例に関連して記述されているが、特許請求の範囲に記載された発明は、このような特定の実施例に不当に限定されないと理解されるべきである。実際に、本発明を実施するための上記方法の各種変形は、分子生物学又は関連分野の当業者において明らかであり、このようなものも特許請求の範囲に含まれるものである。

【配列表】

20

SEQUENCE LISTING

<110> PE CORPORATION (NY)

<120> ISOLATED HUMAN DRUG-METABOLIZING

PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN

DRUG-METABOLIZING PROTEINS,

AND USES THEREOF

<130> CLO00897

<140> PCT/US01/42528

<141> 2001-10-05

<150> 60/241,745

<151> 2000-10-20

<150> 09/739,456

<151> 2000-12-19

<150> 09/818,647

<151> 2001-03-28

40

<150> 09/852,067

<151> 2001-05-10

<160> 4

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 2327

<212> DNA

<213> Homo sapiens

<400> 1

cgcgccctgcc tccctcctccc aggcctgagc tgcctcctccc actgcctttc ctctctcccg 60
cgagtcagaa gcttcgcgag ggcctcagaga ggccttgggg tggcgaccc tacgccagct 120
ccggcgggga gaaagccac cctctcccg cccccaaggaa accgcccgcg ttcggcgctg 180
cgcagagcca tggaaattct cttggctggag acgcgcctgg cgcgccctt ttacctggcg 240
ttcgtgttct gccctggccct gggcctgtct caggccattt agctgtacct gcggaggcag 300
ggcctgtctc gggaccctgc ccccttccca gcgcccccca cccactgggt ccttgggcac 360
cagaagttaa ttccaggaag taacatggag aagcttgagg aaattattga aaaatacctt 420
cgctccctcc ctctctggat tggccctttt caggcatttt tctgtatcta tgaccagac 480
tatgcaaaga cactctgag cagaacagat cccaagctcc ggtacctgca gaaattctca 540
cctccacttc tggaaaaagg actagcggct ctgacggac ccaagtgggt ccagcatcgt 600
cgcttactaa ctcttggatt ccatcttaac atccctgaaag catacattga ggtgatggct 660
cattctgtga aaatgatgtt ggataagtgg gagaagattt gcagcactca ggacacaagc 720
gtggaggctt atgagcacat caactcgtat tctctggata taatcatgaa atgcgctttc 780
agcaaggaga ccaactgcca gacaaacagc acccatgac cttaatgcaa agccatattt 840
gaactcagca aaatcatatt tcaccgcttg tacagtgtgt tgtatcacag tgacataatt 900
ttcaaactca gcccctcagg ctaccgcttc cagaagttaa gccgagtgtt gaatcagtac 960
acagatacaa taatccagga aagaaagaaa tccctccagg ctggggtaaa gcaggataac 1020
actccgaaga ggaagttacca ggattttctg gatattgtcc ttcttgccaa ggaatgaaagt 1080
ggtagcagct tctcagatat tgatgtacac tctgaagiga gcacattctt gtggcagga 1140
catgacacct tggcagcaag catctcctgg atcctttact gccctggctt gaacctgag 1200
catcaagaga gatgccggga ggaggtcagg ggcaccttgg gggatgggtc ttctatcact 1260
tggaccagc tgggtgagat gtcgtacacc acaatgtgca tcaaggagac gtgccgattg 1320

10

20

attcctgcag tcccgtccat ticcagagat ctacagcaagc cacttacctt cccagatgga 1380
 tgcacattgc ctgcaggat caccgtgggt cttagtattt ggggtcttca ccacaacctt 1440
 gcctgcgtct ggaaaaaccc aaagggtctt gaccccttga gggtctctca ggagaattct 1500
 gatcagagac acccctatgc ctacttacca ttctcagctg gatcaaggaa ctgcattggg 1560
 caggagtgtt ccatgatga gttaaaggta accattgctt tgattctgtt ccacttcaga 1620
 gtagctccag accccaccag gctcttact ttccccaacc attttatctt caagcccaag 1680
 aaagggaigt atttgcacct gaagaaactc tctgaatgtt agatctcagg gtacaatgat 1740
 taaacgtact ttgttttctg aagttaaatt tacagctaat gatccaagca gatagaaagg 1800
 gatcaatgta tgggggagg attggagggt gggtggatag ggtctctgt gaagagatcc 1860
 aaaatcattt ctaggtaac agtgtgtcag ctagaictgt ttctatataa cttggggaga 1920
 ttttcagatc ttttcgtta aactttcact actattaatg ctgtatacac caatagactt 1980
 tcatatattt tctgttgtt ttaaaatagt ttccagaatt atgcaagtaa taagtcatg 2040
 tatgtcact gtcaaaaatt cccaacacta gaaaatcag tagaataaaa attttaaatc 2100
 tcaattcact tagccgacat tccatgccct gaccaatctt actgcttttc ctaaaaacag 2160
 aataatttgg tgtgcattct ttccagactt ttctatata ttttatatgt agaatgtag 2220
 caatgtattt gtatagatgt gatcattcct atattgttat tgattttttt cacttaataa 2280
 aaattcacct tatcccttaa aaaaaaaaaa aaaaaaaaaa aaaaaaa 2327

10

<210> 2

<211> 510

<212> PRT

<213> Homo sapiens

20

<400> 2

Met Glu Phe Ser Trp Leu Glu Thr Arg Trp Ala Arg Pro Phe Tyr Leu
 1 5 10 15
 Ala Phe Val Phe Cys Leu Ala Leu Gly Leu Leu Gln Ala Ile Lys Leu
 20 25 30
 Tyr Leu Arg Arg Gln Arg Leu Leu Arg Asp Leu Arg Pro Phe Pro Ala

35 40 45
Pro Pro Thr His Trp Phe Leu Gly His Gln Lys Phe Ile Gln Asp Asp
50 55 60
Asn Met Glu Lys Leu Glu Glu Ile Ile Glu Lys Tyr Pro Arg Ala Phe
65 70 75 80
Pro Phe Trp Ile Gly Pro Phe Gln Ala Phe Phe Cys Ile Tyr Asp Pro
85 90 95
Asp Tyr Ala Lys Thr Leu Leu Ser Arg Thr Asp Pro Lys Ser Arg Tyr
100 105 110
Leu Gln Lys Phe Ser Pro Pro Leu Leu Gly Lys Gly Leu Ala Ala Leu
115 120 125
Asp Gly Pro Lys Trp Phe Gln His Arg Arg Leu Leu Thr Pro Gly Phe
130 135 140
His Phe Asn Ile Leu Lys Ala Tyr Ile Glu Val Met Ala His Ser Val
145 150 155 160
Lys Met Met Leu Asp Lys Trp Glu Lys Ile Cys Ser Thr Gln Asp Thr
165 170 175
Ser Val Glu Val Tyr Glu His Ile Asn Ser Met Ser Leu Asp Ile Ile
180 185 190
Met Lys Cys Ala Phe Ser Lys Glu Thr Asn Cys Gln Thr Asn Ser Thr
195 200 205
His Asp Pro Tyr Ala Lys Ala Ile Phe Glu Leu Ser Lys Ile Ile Phe
210 215 220
His Arg Leu Tyr Ser Leu Leu Tyr His Ser Asp Ile Ile Phe Lys Leu
225 230 235 240
Ser Pro Gln Gly Tyr Arg Phe Gln Lys Leu Ser Arg Val Leu Asn Gln
245 250 255
Tyr Thr Asp Thr Ile Ile Gln Glu Arg Lys Lys Ser Leu Gln Ala Gly
260 265 270

10

20

Val Lys Gln Asp Asn Thr Pro Lys Arg Lys Tyr Gln Asp Phe Leu Asp
275 280 285
Ile Val Leu Ser Ala Lys Asp Glu Ser Gly Ser Ser Phe Ser Asp Ile
290 295 300
Asp Val His Ser Glu Val Ser Thr Phe Leu Leu Ala Gly His Asp Thr
305 310 315 320
Leu Ala Ala Ser Ile Ser Trp Ile Leu Tyr Cys Leu Ala Leu Asn Pro
325 330 335
Glu His Gln Glu Arg Cys Arg Glu Glu Val Arg Gly Ile Leu Gly Asp
340 345 350
Gly Ser Ser Ile Thr Trp Asp Gln Leu Gly Glu Met Ser Tyr Thr Thr
355 360 365
Met Cys Ile Lys Glu Thr Cys Arg Leu Ile Pro Ala Val Pro Ser Ile
370 375 380
Ser Arg Asp Leu Ser Lys Pro Leu Thr Phe Pro Asp Gly Cys Thr Leu
385 390 395 400
Pro Ala Gly Ile Thr Val Val Leu Ser Ile Trp Gly Leu His His Asn
405 410 415
Pro Ala Ala Val Trp Lys Asn Pro Lys Val Phe Asp Pro Leu Arg Phe
420 425 430
Ser Gln Glu Asn Ser Asp Gln Arg His Pro Tyr Ala Tyr Leu Pro Phe
435 440 445
Ser Ala Gly Ser Arg Asn Cys Ile Gly Gln Glu Phe Ala Met Ile Glu
450 455 460
Leu Lys Val Thr Ile Ala Leu Ile Leu Leu His Phe Arg Val Thr Pro
465 470 475 480
Asp Pro Thr Arg Pro Leu Thr Phe Pro Asn His Phe Ile Leu Lys Pro
485 490 495
Lys Asn Gly Met Tyr Leu His Leu Lys Lys Leu Ser Glu Cys

10

20

500

505

510

<210> 3

<211> 31208

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<222> 4027-4221; 8990-9744;and 12103-13769

<223> n = A,T,C or G

10

<400> 3

ccagcccttc ttaggcctct aaatatagtg caaaaagttc cagagttcct ttgttaccca 60
tgaaagcaca tggaaaggctg ctggacagg gcaactggcc ctggagcaga ggagtaactg 120
catagaactg tccaagcttc agaggagtc acaccaccag caagaacctg ggtgggagta 180
ggtgagccaa ggggttccca ggcctcgacc ctgccaagag aactcattag aaggtcacca 240
accacacata ctattcctcg gtcctcatgaa gaaccagg gaccggaccag gcaagataac 300
acaaagctga agtttcagct ctggggcaga gcatggatct gaggtctttg gccctaccac 360
catgcgaica tatgaggggc atcatacaac catcatgatt tgggggagga atagggcata 420
gagggaatcat atgaaaagct gaaatgccat gagttacca gaagaagctg tgaagccag 480
aggattctga gacctgtca aataacaaca tctagtigaa ggttggagtt aggtaggagg 540
taggggaagtc tgggaaagaa ggagctgaaa cacttgctgt gtgtggctta atggaacatg 600
caaggggcca ggacgaactt ggtccagatg aagtcaccac cccctggggc ctgtcttttt 660
tttttttttt tttttttttt tgagacggag tctcactctg tcaccaggct ggagtcagat 720
ggcgcgatct cggctcactg caatctttgc ctctcgggtt caagcgattc tctgcctca 780
gcctcctgag tagctgggat tacaggcgcg cgccaccacg cccagctaatt ttagtactg 840
ttagtagaga tgggtttca ccatcttggc caggatggc ttgateccit gacctctga 900
tcgccccgcc tcggcctccc aaattgctgg gattacaggc gtgagccacc gcgccccgcc 960

20

ccctggagcc tgccttaac acttaccgc caaataaaat ctggctccag agagtggagc 1020
gtaggcttaa ggaattggg gcggaaggc ggggaaggig ggggaggac agtgatagg 1080
agaacagga attgtagcag aaattgggt tatgttcag agctgtcaat gaacacttaa 1140
catatgcctg tcttagccta aatcaatgaa taaatgaatg aataaataaa tgaatgaaat 1200
gtgggcaatg cctataaaga ttgctgggac agggaggig ggggagacac cagcttggg 1260
agtcaggcct gttagatcct agttaccac ctgatacgtt acaaatacta aaaccatcac 1320
tttcaaatta ttttactac attttctgt tatctgtact cgagtttatt tatgtttctg 1380
gcatctagag tcagcccttc atgggcatga gacccaagca gccacacgag gctctgaacc 1440
cagaagagca tatgtcggg ttaattggct gtcacttag aattgttaat aaagttttta 1500
tcccgcatit tcatittgca ctgagattca taaattatat agcaggccct gactgtacct 1560
gtatagtgga attactatat gatggtagc tactgtgcat atcttccccg tttagtttc 1620
agtgccctcg tatcgagc tgaactagc tcatggtaga cgttgggaaat cagggtggg 1680
atcagtigta aaccatttac cggaacacca ctaggcaggc cacaggataa aggaataatg 1740
atggtagacc tccccctacc tctaccacct ggaattttg gtgaatgcc agaattgaaa 1800
agaaaatctc ttgcatagcc atttataatt tggataagg aagaaaaaca atgacctcag 1860
ctttagcatt attttacaat ataaattcag atcccgtagc tgaaaactgt tggactttaa 1920
agaggagcct ccaggagcgc aaaagcagtt gggccgaacg aagcgtgcgc gctttggtaa 1980
ccggctagaa atcccgacgc cgcgcctgcc tctctcccc aggcctgagc tgccttccc 2040
actgcttcc cttcttcccg cgagtcagaa gcttcgcgag gggccagaga ggcgtggg 2100
gtgggagacc ctacgccagc tccggcggg agaaagccca cctctcccc cgcccatga 2160
aaccgcccgc gttcgcgct gcgcagagcc atggaattct cctggctgga gacgcgctg 2220
gcgcggccct ttacctggc gttcgtgtc tgcctggccc tgggctgct gcaggccatt 2280
aagctgtacc tgcggaggca gcggctgctg cgggaccigc gccccitccc agcgcccccc 2340
accactggt tcttgggca ccagaaggta aatgggaagg aaaaaggnta gaaaaggagg 2400
aagagggggc cggaggagga tgcggcagag gagcccagcc ggcagagaga cgcagcttc 2460
ttccatccct ggggacctc cggcttgac cggccttccc agcccgccct gtggctctta 2520
gcatcatttt tcttgcctt ggagaattgc ttcccgag cccacagg aaaggacaca 2580
aaagaggaag ctttggggc tggagagag ctatttaaag aacctgaata tggaaaaaga 2640
aagcgagctg taactcaagt ctgtcttca ttgttcacc aagccttcca catgtgttc 2700

10

20

tttaaaaata gcaigtattt ctaaataact tattagttgc agaaaatag caaaatctat 2760
cccaatcggt ggcacccctta giccatitaa acaagagaaa attttctttt cctaagattc 2820
ttgtgaagta aggagcagcc ccagccagcc actcgagaaa tactgattga tggaaatttg 2880
taaaggaga cgttagctt ttggctctc ccgttttta aatccactcc caccctaat 2940
taaggttttt attcattcaa ccgactctga gtggcaattg tgtgataggt actaagattt 3000
caaagagaag ctaagtcctt cccctgcacc acccaagta ggtagcagct taggcccacag 3060
agagaaaatg aaaatttaag gcaatgggtg ctttactaga ggcctagaga caagggaata 3120
tcgtcggag gaaagtatac atctccgctt agagaaggaa ggaagcttg tgaaggctg 3180
agcagagctt taaaggatgg ttgggtgggt tgggaaggc attccagcag agctactaca 3240
cgatccittg gttcccccac ttcttagctt ttcttatata aagcaaccac ttcaactct 3300
tttatcggtt tcttcggta tttaaatact tatttgtaa atagtattac catattgcat 3360
ctattaatit aataagttaa gacatctgtt gtggtttaga taiggtttgt tcgtcccccac 3420
caagcccat gtgaaattt gattcccaat gtggagggtg ggaatctatg ggagatctt 3480
gggtcattgg gatggatccc tcatgaatgt ctgggtgcag ctgtctctt cataagttct 3540
cactctctta gicccctctt aacccccaga actgattgtt gaaaagagcc tgccacctcc 3600
tccctctctt ctccctgtt ctccatgtt ggctctgca cacaactgct cctgttact 3660
tccactatga ggggaagcag tctgagatcc tccgcagatg cagatgcaa tgccatgctt 3720
ctgttacagc ctgcagaatt gtaacccaaa taatctctt tgtgaatgac ccagccctag 3780
gtattcttt acagcaacac aaatgtacta agacaacatc cacctatgaa ctctttatg 3840
acaggcaatc acttacactt catattccac tgcaccagta actatatagt attgtattt 3900
ttaaatagaa aaacttctat ttgtattatt ttattatgc aaatgttatt tactgtgat 3960
ctaatggct ccttttctt ttatttctt ttctcataga actttttccc cacccccaca 4020
gtattggnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 4080
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 4140
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 4200
nnnnnnnnnn nnnnnnnnnn ntgttatgta tctctactgt ctcatgaata ctatgtctc 4260
tgtttttta atgaattgt ttggcatcc tgtcaaaaa tcaattgacc ataatgtca 4320
aggctattt ctgagcttc aattctaac catgatcta taigtctat ctaactcatg 4380
gacacagaga gtagaaggat ggttaccaa ggctgggaag gatagaggg agctgggga 4440

10

20

ggaggtaggg aaggtaatg ggtacaaaa aaatagaaag aatgaataac acctactatt 4500
tgatagcata gcaggtaggc tatagtcaat aataacigta cacittitaa taaagagtgt 4560
aataggattg ttigcaactc aatggataaa tgcitgaggg gatgggtacc ccattcttca 4620
tgaigigcct attcacatt gcatgcctgt atcaaaaaa tctcatttac tccataaata 4680
tatacaccta ctatgtatcc acaagtatta aaaattataa ataaataaat tatatagcta 4740
tccttatgtc agtaccacac tgccttactg ttgctttgta gtaagctttg aaatcaggaa 4800
glatgagicc ccgcacitt ggtattttcc aagattattt tggctgtttg gaatccttga 4860
tttctatata aattttagac tcagcctatc aatttctaca aggaaccag ctagggttct 4920
gcttgggatt gcacigaatc tgtagatcag ttggggatt attgccatct taagaatatt 4980
aggctttctg atccaigaac acagaaagcc ttcccgltta gtiaggatcat cttaatttt 5040
ttttgttgtt ttttttgtt ttttgagaca ggtcctctgt ctgtcgccca ggctggagtg 5100
cagtgacgca atctggctc actgcaacct ccgcccctcg gattcaagcg attctcctgc 5160
ctcagccicc caagcagctg ggactacagg cacatgccac cacaccaact aatttttgta 5220
ttttcagtag agacgggtt tcaccatatt ggccaggcta gtctcgaact cctgacctcg 5280
tgatccacc gcctcacct cccaaagtgc tggattaca ggctgagcc accactcccg 5340
gctttcttta attttttta acgatgttt tgtattttc aaagtataca tcttgcatit 5400
cttttgttaa atttattgt tttgttctt ttaatttcat ttcagactat ttattgcatt 5460
catagtgttt tagagtcac attcctctt gactgtcact aagtttttt tttctgttt 5520
ttgagagggt tctatcagaa ttttgcatg cagagatgac ggacatgtca aactgtctaa 5580
tattaccaac cctcccat tctcagatca ggatccttt ggtagttac catgcaggga 5640
aatctagtat ctaaggctca aaaggtaga cgttttaca taggcagtaa catttattg 5700
ctacataata actacatatt tatggagtac ctgtgatatt ttgatacgtg catacaatgt 5760
gcagtatca aatcagggtg tttaggtat tcatcattc taacatttat tatttattg 5820
tgtttggaac attcaagtc tcttcaagct cticagaaat attcaatata ttattgttaa 5880
cagtgctatt gaacactgga acttattcct tctatctaaa gacagtaaca ttttaagtat 5940
agtcataagg ttacagaagg ataaagtgtg tatagggaat attccttaca agatgagaat 6000
ttcattcctt actcttagta atacaggct tcaaacaatg caaggatatt cctcccttgg 6060
agctttgaac atgcagctc gtgggtatat tgcctcctt gcaaattatt cctaaaagag 6120
gcttgcctg accattcaga ctaaaatagc acccttagta ctctctatct ccaacctat 6180

10

20

tattattatc ttggccctta tcactctctg acactatact gtatactctt ttgcttgttc 6240
gtttattatc caccactaac tacaatataa aatcigtigag aggtaggatc ttgttttgcc 6300
actataaacc tagtgcattg tacagtictt ggtgcataat aggtgctcaa taaatccitt 6360
gttgaatgca taaatataat aggtgctgag aaaatttatt tattcaaaga tcaatttact 6420
gcatagaata ggccaggtag tttagacattt attcaatagc caacatatgg gacctaggat 6480
gtacatatgc aagtgtgtgt gtgtatgtgt gtgtgcatct gcatgtgtac ttggtatgtac 6540
tgcagagaac atctatgtag ctaagtagta taaagcactt gggtccaga gttaaactgg 6600
agtttgaatc ctcattagtg gttgccagct gtacacactt gggcagatca ttaacctag 6660
tctgtagggc tcaatttctt catctctaaa gtaggatig taatcatatc tacttcatag 6720
ggttcttgat gtaaatatta aataacatag aacatggaaa gcatttagca gcacctagtt 6780
catagcagtg ctgtataaat gttagctgtt gctatttggg ggcactatgc attttctgaa 6840
cattttctgaa caatgtttac taaatataat tagtaccctt ttcaagtgt atttagatgc 6900
ttctctgggg atgaagaaat ataaattaaa tatagtacag tattcacaac agttttctgt 6960
cctttttgtc tagtcaggag ttacaaaaag tataatgaaa tactttcata tggctgggt 7020
gtttatgaaa attttttacc taaacaaaca attgtcatat tagttttaca tattcatgag 7080
ggcaaaggcc ttgtcttctt tatatttctc tgtatctcta ccacctggtc cgtgtgatag 7140
acaataaata ctgtgtgtt tattgtttgt aaatgaataa atgaaaaat attcacattg 7200
ttgaaaacca ctactctgga tagtcagtgg gtgtttatca ctggcttgat tatggcaaca 7260
ttaacaaaaa agtgcagtat tttagaaact aggtttcaag actctcaacc tttagtggc 7320
cttgaactat ccagagaaca ctttaagggt taaaatgtc aaatgataac agagaaaaat 7380
gggagccaga gtgtccacc tctccagagg atgagagcaa acaatcctgc agcagatacc 7440
gtgtgattgg tcacacgagg aaaaatctgg cagccttaag attactttgc agcggggac 7500
tcccaccatc atgtcaagt gtgtagatgg gcacaccaa acacacacat gcaggtgccc 7560
tccactttac acaagaagca aatgtaaatg aatctgttt tcagtgttt agagaaacaa 7620
tttaagttag ccattactca tctgttcta aaagcaaaaa ctcttctct ggtagtagta 7680
tttgactct cattgtaaa ttttggagc tgaagtttt gtatttgagt ttgctttaag 7740
attcacacat ctgtgtaaat ggacctctg ttgttgggg gagaatttgg attttcttta 7800
tagatagagt tggcaatttt tttagagaga gcatttactg ctaagtcag agaaataatc 7860
actggtgcat aattagagag aggaacagga agaagaaatg gtgagctgga tgtagggtca 7920

10

20

tgccccattt agtaacigtg agtttccac ataggaaata ctcttttta gcttccagat 7980
cccactccaa tctgagtgig tgaigtggc aagtgaaggca gagagtgiga ctgggtcac 8040
ccctatagg gacaagagtt cacagttaa gtcatcaac agtgacttgg tctggggta 8100
caggatata taatatigag aagataaata cactaacit gtttagagaa ttatcccca 8160
agcittagaag tcccaaagaa agcatgttat gtcacttcca gaaaagtc aggtccct 8220
gcttgigtga cttatcagg tctgaactc agcttgigtc tataagagg gacaggcca 8280
gcttggctgg ctaattact ttacttttt cactgcagt taticaggat gataacatgg 8340
agaagctiga ggaaattatt gaaaaatac ctctgacct cctttctgg attggccct 8400
ttcaggcatt ttctgtatc taigaccag actatgcaa gacacttcg agcagaacag 8460
gtaagaagag gggaaagct ctgggaccta ttcttctag aagtgaatg cataaaacc 8520
ataggcaaga ttccaaagca aagattgggt tgggcccit aagagacaca gcagcaagta 8580
tgggaggig acaggttcc taccaatact gaagggtt cccatattct cccagtcct 8640
ttgtctggt caggtatga tgggcacgt gaagtcgta taacttaaag cctagctggc 8700
attaccagac ttgccaggca aggttccct tggcctctgt ggttttatg acttcagtg 8760
cagcaacact tccactcct accctggc tggagcataa gtctcaagag ggtgggaaat 8820
cagcagtaac tctacctg ctgggtcagt atgaaagcct gaatgctaga tcattaatt 8880
acccatcaga cctctgatn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 8940
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 9000
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 9060
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 9120
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 9180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 9240
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 9300
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 9360
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 9420
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 9480
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 9540
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 9600
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 9660

10

20

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 9720
nnnnnnnnnn nnnnnnnnnn nnnntctgct tgactctgca gatcccaagt cccagtacct 9780
gcagaaattc tcaccicccac ttcttgggat gtatgtgcaa atgagaggta taaccacctc 9840
tcattcaaag tccccitcc atagtagagc atgccaaaga aacigaaac tgaattcaaa 9900
agcacaaga gtgcaaggta gagctatact gaacgttatc ta~~gggg~~aaag attgaagg~~gg~~ 9960
agctctaagg tcaacacacc accacitccc agaaagctc ttcattcgtt tctctccac 10020
aaagtcttat tctcaaggca gcagatacat gaatctgtcc cctctctctt taaaactaca 10080
gccttggcca ggcacagtga ctcatgcatg taatcccagc actttgggag gccaaaggtag 10140
gaggatcact tgaggicaag atticaagac cagctgggccc aacatggiga aatcccatct 10200
ctactaaaaa taaaaaait agccaggcat ggtagcatgt aggcctgtag tccactact 10260
tgggaggctg agacatgaga atcgcttga cctaggagg~~t~~ ggaggttgcc gtgagctcag 10320
attgtgccac tgcactccag actaggtagc agagcaaac tctgtccgca gcccacaaca 10380
acaaaaaaaa aactaccaa actgcagctc caccatccct attcttgttt tctttatcct 10440
tctctcgttt tcttggatgt ttctcttct ttttggagt cctttatttc cacatgcgag 10500
tcagtaaaat ttgtctctag agtttggcaa tattctgtca gcagataaac taagctcttt 10560
aattacataa ttggtattta ttttaacaa gacatgaatg aaagaaaaga atataggctt 10620
gtattaggaa ccacttaaat ttgaatcttg cccctctctg cattgactag ttaaataatga 10680
tcttggggaa gtcatttaat ctcctccat ctcagtttcc tcatcttga caataaggat 10740
gagactcaca ttgttgggct gttatgagga ttaaataaaa tacatatitt tagcactaca 10800
tgtaatggcc accattgtat gagtagcaga tcatgcatca tgagcctgga atgttgttaag 10860
cattcaatga atggtatcaa ttaigtatta ataaacttta aagtcctttt aaagccaaat 10920
cctaataacc agtctggcaa tagaagattg tgaagcatta gccttggtaa gtattccac 10980
atagtatcat tcatagacct gggctcaagg aggaataatc aggggacaga gtggacactc 11040
ttgtctcttt ccttgtgaat ttaigtatcat catatagttt atggattggt ttggagtgga 11100
aaggaaatca ctgtctctgt tactagtgtg agctagggag taggttggct accttatgta 11160
ttcactttca gttaacctcc acagcaacac agggaaaaag gtatttagta tcatagtcca 11220
ttattgagaa aagtaaacct caggaagatt gactcactta ttcatgtact acataggtag 11280
taactggiga tttaggatt agcgtgctaa tcttataagg ctttgaaatt tattagactt 11340
tgaaactgtt tctcacaata ttaaatacat ccatcccaga ggtaagcttc taaattcacc 11400

10

20

ttcatctatt aaattgcatt gcacattaat acgagtacta ctttgatact ccactgttgc 11460
atgactgcct gtgggtcatg gttactccac gctgccctgig ttcttcatct atccctcatc 11520
tcatctaatt aaatggcata aggttttctg ccttttatit ctcaaggaaa aggactagcg 11580
gccttagacg gacccaagtg gttccagcat cgtcgccctac taactcctgg attccatttt 11640
aacatcctga aagcatacat tgaggtagtg gctcattctg tgaaaatgat gctggtaagt 11700
aaagggggaa agtgctctgt gcatigcgaa atgctcccag caatggacag tattaggtat 11760
gtgttttgig ggccatgaaa ataaaaaatc agtttctaaa aatttaacca atgtacacgt 11820
acttatigaa caatagggtg ctgtaaaaaa ttgttatgt tctttgagtg ataataataa 11880
taaaaagatc tggctctctg tcttagatat attttgagat tttatggcag caaaccaagt 11940
accaaattgt gatagttaga tagtaagtc tglagaigig tticatggag ggccggctctg 12000
tacaaaccia ccccaaagtc tgaggaaact gagaggctga agaaaaaggc tgacagtttc 12060
ttaaaaagaa acattcaata gaggccttca aacaaaaacc atnnnnnnnn nnnnnnnnnn 12120
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 12180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 12240
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 12300
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 12360
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 12420
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 12480
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 12540
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 12600
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 12660
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 12720
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 12780
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 12840
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 12900
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 12960
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 13020
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 13080
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 13140

10

20

13200
13260
13320
13380
13440
13500
13560
13620
13680
13740
13800
13860
13920
13980
14040
14100
14160
14220
14280
14340
14400
14460
14520
14580
14640
14700
14760
14820
14880

10

20

actatitita ttatggacaa ttattattaa tacaatatata agtaggcact taagagticc 14940
agacatacat ggaatatggc tttttgcaca gcgattgcag taataataat gacaagctaa 15000
aaacattcat gcaacatagg aatggagagt ggaacagagt aaacatggac atgcacccga 15060
aagaatatig attcaaaaac agtttttagca agcataaaca caaaagtiga aatagattaa 15120
gcittitaaag caattcaaca ttacttgtca tgaatgccat aatggagaat acttatcaag 15180
cagigaatta atccitcatic agcttcacca ctactagca gtactagta agttacttac 15240
tgctttgitt cagtgicatic tataaaatgg agattaaaaa agaacctatic tcatacattt 15300
gttgttacga tgagtgggtt aatatatata aagcatttag gacagtgcct ggcactgaat 15360
agaigttaaa tglaaagiat agttatgtca aatgtctttg ctccaggaa ttttgcaaga 15420
cacaccaaca tatgcacact tacacataca tataigcata catgcacata gatattataa 15480
agaggacact cagagaagca ggtataaac aatttaaggc ataaatgggc attataaata 15540
gcagcagtic ccaagtcitt ctgcattcatt gcacacacag aaaatgttaa tgtttttgtg 15600
cttcattgga gtaaacagga atggatttgg gggaaagctat acagaacttt gtaaaaaaaa 15660
atctttactt tttaaataat atacaattat gatgaaaaag caaaatgcaa agtgttaggg 15720
aaaatattaa atgttaaatt tattcaaac ttaaaacctt ttcaattttt tttttttttt 15780
ttttttgaga tggagtcctt atcactcagg ctggagcgca gtgggtgtgat ctgagctcac 15840
tacaacctcc acctcccagg ttccaggcaat tctcctacct cagccttctg agtagctggg 15900
attacaggca ctgccaccac acctggctaa tttttttaaa ttgtttattt ttatttagtc 15960
aaatataca atattttatt ttattgcatic tggattttta gtaatcaca aaagccattc 16020
tctattccag ggtttctcaa ccttcagcac taatggcttc tttagattaga taagtccttg 16080
ttgtcaagat gtgtgcattg taggatgttt agctacatcc ctgacatcta cccactcgat 16140
gtagtagagc tctgatagtt atagcaacca taaataactc cagacattat tgaatgticc 16200
caggccccc agttgagaac cactgccctg taccagggtt gttagagaaa ttatttatgt 16260
ttcttgttag tacttgtata atttcattat ttcatattt aaatcagaga tctaaactcc 16320
atttagaatt tattcctata tatgggtgtga ggtattgatic taatttttcc aaatgtttat 16380
ccagtgticc catcaccatt atttaaaagt ttatcttttc aagtatttg agataacat 16440
cacattctaa acggatacat gtactgggtat ctgttttggg taagagtata ttggatgtt 16500
ctcgigtatt ccatgtatct atctaccaat gtaccagaat cacactgttt taattaagga 16560
gattttgtgg cttttttcaa cattaataga cttattttt agaaaagttt taggtttgca 16620

10

20

gaaaaattca gcagaaagta cagagagttc tcatattacc catgtaacaa acctgtacat 16680
gtacccctgt atctaaaata aaagttgaaa ttttttaaat agtaaaataa tattacctct 16740
gttccatatt ttgttttgt tttttttc tcagctcctt caattataaa tatattggca 16800
tttcttggc tgtcttctat ttcatccat ttatittaat aacttttccg tgaagataaa 16860
atattagact gaggaagaaa agaataattg gtcacttgca tctaaacttg aaatcatctt 16920
aatittattg cccacatact gatggaaact atgttttta ttgtgttgt ttatctttgg 16980
agctttaatc aaaagtcctt ttgatgagaa aataaacctt ctgtgaaaat tagatctatt 17040
taaacgtctg gaaatcaggc aagattigaa gctattcact aacctggct tgccttataa 17100
tttatttgac ttggccatca ctttggtaat tggaaactat tttctaccc agatacaata 17160
atccaggaaa gaaagaaatc cctccaggct ggggtaaagc aggataacac tccgaagagg 17220
aagttaccagg atttcttggg tattgtcctt tctgccaagg taaatcttct aaatttctaa 17280
gccigtctca gtgaccagtt aattaigtga gtaggtgggt aagtgaggat gggatgggga 17340
gacaagaata aaaccgatg actaaattta actgtacttt gaattgatga gcagcttcat 17400
gcaatttgag acaaagagag aattctgcaa ctgtgtcgt agaggagggt tagtaaagac 17460
taaacgaacg atttgacaag atttgaggat tgcataatgg atacatggat tttagggcat 17520
catgaaaaaa tggtcacatg gataaacgta aaaattatga tgataaggtc ctgggaaatc 17580
tgggagtttg aagagaattt ctaggccctg ttgatcgagg gccctttgtg caaggcctgc 17640
tttcttctc taaccttggg tctcctttat gctttgggca gaatatgggt tataccacat 17700
atttgttgaa ctgaattaaa atttaaacc cttattaaag ctctgatttt tcccccaaa 17760
tcattattgt ggttgtatct ccaaacttt ataaacggc attttattta aaatatttgt 17820
attgtacttt ctaggatgaa agtggtagca gcttctcaga tattgatga cactctgaag 17880
tgagcacatt ccgttggca ggacatgaca ccttggcagc aagcatctcc tggatccttt 17940
actgcctggc tctgaacctt gagcatcaag agagatgccg ggaggaggtc aggggcatcc 18000
tggggatgg gcttctatc acttggtaag atctgcaccc cttaaatttc ctgctagttt 18060
tccccctgag attttgcttt attttttgcg ctggtacctt agtgacccta gtgcctcagg 18120
atatgtgtag gtgaaacaga agaagtaggc tactttctg ttctttctaa agagagctcc 18180
aaattattct ctgtcttct aggaaaaaaa aaaaagtta ttatccata aattgtctgt 18240
cattggtttt ctaatcaatg gtgtgtgaaa tgtcttattt ctttatttca ccttggctct 18300
gatgcattgg aatgaggac ttgatccctg ggctggcact tagaacttaa acaatagggt 18360

10

20

ccaagtggag ctcccttctt gagagagctg aatgattagc tgcattattt aaggctcatt 18420
ttagacatct cccagccgct tgtaccaat ttatctctc aggattgatt ttagacttca 18480
gacataatat tcatgatat atactatagt taagttagc aaatatggac tgaggacatt 18540
ttaaatactg agacttttt ttgactaca atttatgtg ggccctgtct tgggtgagct 18600
aatggictaa tacaggagac aggagacaga cctccaaatt gcagtgtagc ataatgaggg 18660
caatgataga gatatgtgct ggctaacaca aagacataga agacaggtag ctaccctggc 18720
atgggagctc aaggagactt ccttgacatt tacgtgact gcaggataag taggagttag 18780
ccaggtagaa actgicactt ctatcttgc agactttaag catatactgc tgttaataaa 18840
gcccaggita tgcgttttgc aaagataaaa tgggttctg acataatact ggtcaaggg 18900
acagaaagac agaaatgcta aggacaattc agcagcagac cagataaaaa acaccatatt 18960
tcatatgcaa aagtcaactc aattgaaaca ttgtaaaac caaatttgac attataaaag 19020
tataicagag atctcatitt ataaggaaat agaagccctt tctaccata aactaaagat 19080
ttaatctata tagcacaana tacaatgtg agtaatcatt ttttaattat ttttaactg 19140
acaaaaatg tgcataata tttatata atatgtatgt ggtatata atatgatga 19200
caacatgata tttgatata tttataact gtggaatgac taaatctatc aatggacatg 19260
ttcatiaact catactatc attttttgt ggttaaggaca tttaaaatc accctcttag 19320
caattttcaa gtatacaaat ttttagtaac tccaatcaca tattgtacaa tgcattctct 19380
aaacttatgc ctctgtctg actgaaattt tttatcttt gactaacatc cctgtaatcc 19440
cccattctcc cacagccctt ggttaaccact gtctactct ctgctcttt gagtttaatg 19500
ttttagattt ccacatgiga gatcatgtg aattgtctt tctgtgctg gcttatttca 19560
cttagcataa tgcatacaa attatctct gtgtcataa atgacaagat attgtcttt 19620
tctatggcta atgttagtc catgtttat atatatacca tgtttcttt atccatttat 19680
ccagtgatgg acacttaagt tgatttctat atctggcta ttgtgaataa tgcgtcaatg 19740
aacatgggaa tttgatgtc tcttcaatgc actgatttca tttcgtttg ttttatatcc 19800
agaagtggaa ttgctgcatc atatggtagt tctattttta attttttag gaaactccgt 19860
acaattttcc atatggctgt actaatttac attccaacca aaagtgtata agggttctgt 19920
ttctccaca tctcaccaa cattgtctt ttggtaata accattctaa tgagcatgag 19980
gtgatgtctc attatgggtt taatttacgt tttccgtatg attagtatg ttgagcatg 20040
ttttaaatc ctgctggcca ttcattgtct cttgttagga atgttatitt aggttttct 20100

10

20

catitttaaa tctagttatt tgttttcttg cttttgaatt gttgtagttc ctcatatatt 20160
ttgaatatta accccttate agatgtatca ttgcagaca tttctccca tccitttaagt 20220
tgtctcttca ctatgttgat tgtttccitt gtgtgcaga agctttttag ttgtctgcaa 20280
aaccatttat ctattttttc ttctgttgac tatacttcca gagttgtatc caaaaaatca 20340
ttgccaagaa taatatcaag aagcttttct ctatgttttt ttctagtagt ttatatagtt 20400
caggtcatat gtttaaatct ttaatccatt tttagttgat ttttgtatat ggagtgagat 20460
aaaggtccac ttttatttct ctactagtgc atatccagtt ttctcaacac cattatttga 20520
agatactgcc ctttcaccac tgtatgttac tggaaaccttt gtagatcagt tgacaataaa 20580
tgtgtgggtg tatttcttga ctttttatcc tgttttatta gtttataigt cttttttttt 20640
agaagctcia tgcgtttttg gtgactagag cttgttagtc aatttcagat caggtagtat 20700
gatgcactcc agctttgtct tttttgtcca aaattgcttt ggctatttga gtttttttat 20760
tccatacgaa ttttagggct tttttttttt ttcgattact gigaataatg ccatttggat 20820
tttgatggag attgcatlga atctttgggt agtatggata ttttaacagt attaatgctt 20880
ccaattaatg aacacagggg attttgcaat ttgtgttttc ttcaatttct ttaccagtg 20940
tttttttctt aatttaattg ttttatttcc atagggtttg ggttaacaggt ggtgttttgt 21000
tatgagtaag ttcttttagt gtgatttgtg agattttgat gcacccatca cctaagcagt 21060
atacacigia cccaatttgt agtcttgtat ccttcacctc cctcccacca tttcccccaa 21120
gtccccaaag tccatttgtat cattcttaig cctttgcate ctcatagctt agtccctact 21180
tatgagttag aacatataat gtttgggtct ccatttctga gttacttcat ttagaatatt 21240
ggcttccaat tccatccaga ttgttgcgaa tgcctttatt ttgttccitt tcatggctga 21300
gtagtattcc atagtatata catcccacaa tttctttatc cattcttgat tgatgggcat 21360
ttggactggg tccatgtctt tacaattgcg aattgtgtct ctacaaacat gcaggtgcaa 21420
gtgtcttttt catataatga cttctcttcc tctgggtaga taccctgtag tgggatttgt 21480
ggatcaaatg gtagttctac ttttagttct ttaaggaatc tccacactgt ttccatagt 21540
ggttgtacta gtttacattc ccaccaacag tgtagaagtg ttccctgttc actgtatcca 21600
caccatcatc tattattatt tgattttttg attatggcca ttcttgcagg agtaagggtgg 21660
tattgcactg tggttttgat ttgcatttcc ctgatcatia gtgatgttga gcatttttct 21720
atatatttgt tggccatttg tacatcttct tttagagaatt gttattcat gtctttgtc 21780
cattttttga tgggattatt tgttttttct ttgctaattt gagttccctg tagattctgg 21840

10

20

atattagacc ttgttggat gtgtagggtg tgaagatttt ctccactct tgggttgc 21900
tgtttactct gctgattatt tcttttgcig tgcagaaact ttttagttta attaatccc 21960
acctattat cttttcgttg ttgttgtttt tgggttgc ttgttttgg ctgtttttg 22020
catctgcttt tgggttcttg gtcataagt ctttgcctaa gccaatatct agaaggggtt 22080
ttctgaigt ctagaatttt tatggttcag gtcttagatt taagtccttg atccatcttg 22140
agttgatttt tgtataagg gagagatgag gatccagtt catgcttcta catgtggctt 22200
gccaatatc ccagtacaat ttgttgaata ggttaatat ttaaagcttt atatatattag 22260
gtgttctat ttgggttaca tatttatita caactatcat atcttctga tggattgacc 22320
cctttctcat tatataatgg tcttcttgc tctttttaca gtttttgc taaagcctaa 22380
ttgtctgat aaaagttcag ctaccittgc tcttittgg ttctatttg catggaatat 22440
ttttttcaa ccttctcat tcatctatg tgttcttcta aagatgaaat gagaigctgt 22500
aggggcata gcttgggtct tgttttattc attcattcag ccacccttt gattagagaa 22560
tttaattcat ttgtattca ggttaattat gacagacaag gacttactac tgccattttg 22620
ttaattgttt tcttgatgt ttatagatct ttgttctt tcatctctc ttactctttt 22680
ccttttgat taggtgcttt tctctagtgg tgtacttga tttttacttt ttatcttttg 22740
ttgtctact ataggtttt gctttgtgg taccatgagg gttacataaa gcatagtatt 22800
aaaaggctat tttaaactga taacagctta acttcaaca cttaaaaaa ctatacactt 22860
ttactctacc aactgccctc cttttatgt ctttgatgc ataatttacc tagttttgga 22920
gatgttccc cttatttgt atccctaac aaattattgt agcaacagtc atttttaata 22980
gttttggctt ttaactttat actagagata gaatttaata acataccacc actacattat 23040
taggttattc taaattgact atgtatttac ctttatcagt gagatttttg ttccaattt 23100
tcatgttgt aattagtatt ctttcatte aacttggaga attcacatta gcatttttg 23160
taagatgggt ctagtatgg tgaacacct caactttgt ttatctggag atgtctttac 23220
ctctgcttca ttgtgaaata taactttgt tccatgattg aaatggacaa aatgttttt 23280
ttaattatgc aaagtccag ggttaagcaga attactcttt ttttttttt ctgagaccga 23340
gtttcactct tgttcccag gctggagtc agtggcgcaa tctctcagct taccgcaacc 23400
tcgtccccc aggttcaagc gattctctg cctcagctt cctgagtagc tgggattaca 23460
ggcatgcacc accatgctg gctaattttg ctttttagt agagacggg ttctccatg 23520
ttgtcaggc tggcttgaa caccgacct cagatgatcc gccacctag gccctccaaa 23580

10

20

gtgctggat tgcaggig agccactg cgccggcaga attactctta tttatcctga 23640
gcttgaggaa gaaagaattc aaaattaaaa ttccacatta cctaattggc aaagcctgca 23700
ttcaaaataa gtaatcagaa aaacatataa aaacacaata agataaacag actaaatata 23760
tgcagtcatt ttaiggaacc aatctgacta gatiggaigc agactaggta ggatgcaaat 23820
ttaaaaaaaa ctttattctt ctccactta taaactttaa acctgccttg tggagcaagt 23880
tctttttatc tctgggaaa gatcctgagt aagtcctata gaggctcat tcatitaaat 23940
cacaagaaca atcttaggtc agtaattaaa ctatctggcc cagtgtata ctgaaacttt 24000
caaatactta tccactigag ctctctttc catcccagct tggacttct tggctctag 24060
aagccagcag tggtttatca tgcacttatt ctactgact agctcccaa taccagtag 24120
ctgctgtttc tggccctcc aggaatggtt ttaggaggaa aggggataag gaggaaagg 24180
ctggactat tgtatcatg ccaaagggtc tggggatatt tccatgttc ctttctctc 24240
aagaggaaac tccctttctt ggagactctc tactagaac ttccagagg tgattcagg 24300
gacaagagaa taatgtctt taggcagact cttttcaag ctggctccag agctttccct 24360
cttgccagtt aattggttta aggacacagt tgcacatctt tgccttgcct ctgctgctgt 24420
cctctgctt tctgtctgt ctgagttata gcccttcaca tcagtcctgt actcccaaa 24480
ctccaaggag cacaagtcag atcatctaag tgaatctctt gaagcctctt gtttaagatg 24540
ggggaagcac ctttctttt ccatggcact ctggcattcc aacaacactt taaataattt 24600
tttctctcaa aattcttaag cctctctctt ttaatcttc gccattttta tgtattatta 24660
ctttatatga tgagctaaga gttacaaaac tggtttttag aaatctctt agcaaagtgt 24720
ttactgctag tttagcagct cactttataa taaggatata tgatatattt ctltggcttc 24780
tctgctctg ggacctcagc tcatctgag gcagagagtc ccattttaac attctgttac 24840
ataaaccagt ggcaaatgg cttaacctg agggtaataa ttaccaggaa caaacagaaa 24900
acagaaaaaa agtaaacctg ttatgatac tgagtcctt ccttccctca tcttcacagg 24960
gaccagctgg gtgagatgtc gtacaccaca atgtgcatca aggagacgtc ccgattgatt 25020
cctgcagtc cgtccatttc cagagatctc agcaagccac ttaccttcc agatggatgc 25080
acattgctg caggctttta cttcttttc ctaagcagtt cttagaggct atgggactct 25140
ggagaccaca gtgacaaaga ttatgagtc tcttagcact tggagaagtc aaaagataat 25200
gctaacaagt gacttaggtt ttatcaccta tgaggagctc agaggataat gctttgttca 25260
gacatgaatt tcaatgactt tcccaaaggc acatagccag ttgcagcaa gctaagccca 25320

10

20

gaatccatgt ctctggaatc ccagcccagg gtccttcca ttgtggaca tcatttctaa 25380
gataatcttt gtttggctga gtttgagacc gagctgaaac ttcatggaaa atagcaccag 25440
catctttatc tgaaagacca agggggatct ttggcccat catcataata tcacccttat 25500
aaatatacaa catttaatag ttaatataga gccttcagac ccattatctc atttttccc 25560
ttggaatcca atgttaacag atgcttatac aatgalttac agttcactga acacttttaa 25620
gtactttcaa tgtggcccaa aatccagagg cagccccaat gtgtagatga cattaactga 25680
tgtgagcaga gctagaactt gtgcggagac ccigagctcg gagcctagag ttcttcggaa 25740
caacacaggt ttctgagcag ggcttatagg aagcagagg gtcattgag acatattatc 25800
tgattcaatg ttctattaat tcatgtctta ggaagcaagc caacaggatt gcttctggca 25860
aacacctaca gccgttact gtaactttgc tgacagacc agaattaatt tctggaagct 25920
agaattatit ctggaaacca aataaccctc acattctctc tctttgttt tgtactctgt 25980
ttctcccaa accacatgga taittgcaa aattctccac ttcccatatg tgaatagcac 26040
caatggaaat ttgtcatggg atctgcatga cagaatcaca gtctgtgtgt tgtgtgtgt 26100
cgttttctc tcaagacaga gtcttgctat gtgcccagg ctggagtaca gtggcgtaat 26160
ctcggctcac tgcaacctct gcctcccagg ttaagcagt tctctgcct cagcctccc 26220
agtagctggg attacagggt cacaccagc ctggcaaat ttgtatitt tattagagat 26280
gggtttcac catgttggc aggttagct caagctctg atctcgagac cagcctcct 26340
cagcctccca aagcgctggg actacagcca tgagccacig caccagcca gtctgtgt 26400
tttataccia aattgtctc aggtgtgtt aatagtcct taataggtat ttaggccagg 26460
cacagtggc gacgcataa atcccaat ttgtgacac caaggtggga agactgtctg 26520
aagttaggag tctgagacta gcctgggcaa cataggaga cctgtcttt acaaaaaaa 26580
aaaagagaga gatagccagg catgtgtgt catgttgtt ttctgtcta ctgggggac 26640
tgaggcagga ggaacactg agctcagaag ttcaaggta ccgtgagcaa tgttcacgcc 26700
actgtctcc agcctgattg acaggccaga cctgactct aaacaaaaac aaaaaacaaa 26760
tatttaagta atttccaaac atagcagaaa atataagcat ggtttatcac ttgatatga 26820
caccaacagc tacttaagat agagtcata attcagtaaa ttgtgtgtg gaaagctaag 26880
gtgccaacc aagccgcatc ttcttaggt ctcctcactg gtgtcatcag ctacagcagg 26940
cagagcatg ccaggagcta gctcttccct tcaagaacaa aagcttgtt taagagcaca 27000
gtagcccaca actgtctct tctctgcag tctctttat tccctctt tcttagggat 27060

10

20

caccgtaggt cttagtattt ggggtcttca ccacaacctt gctgcttga aaaacccaaa 27120
ggtagtattc tctctgtac ataaatactt ccaagaacta atgctgtgca agtcactttt 27180
tggtagctaa gcacagaagt ggctatataa ttaagggaaa tgacacaaat taaacaaaaa 27240
taaacataaa agccaaaaga aatgtaaaac tttctatgt tcttgaaaca ctttgacgt 27300
gtatcagta tttcttcat gtaagccact aagggttaag atctattact tgaacagga 27360
agctggagta tatgtctgt taataattgg ccacatcatc attttgactt gatttctaag 27420
tggaigcaca tccatticta agtggatgta tctccatagt gaaaataata ccacttgcca 27480
tagtattttt gtttgccctg giatcagaca aatcagctgt gaagctgcaa ggctgcagg 27540
tctgaaggta cactgcccag tgtagttagc acggccaca tacggctact gagcacaiga 27600
catgtggcca gtgggaattg agttgtgtg taagttaaa atacgtgtgt gattttgaag 27660
acatagtacc ctaaaaaaat gtgaaacatt tcttttagt aattatttat attgattaca 27720
ggttgggaatg gtaatttttg gttaaataaa cttctattaag attaacttca ctttttaaaa 27780
atgtgaccac cagaacattt taaattacac atgtagatca cattatattt ctatgatcg 27840
gtgctagggt gtaggtgaag aaatgtgtc atgtgtttg gggatgggt tgggttgt 27900
cctctcattt caggctttg accccttgag gttctctcag gagaattctg atcagagaca 27960
cccctatgcc tacttaccat tctcagctgg atcaaggiga gaacaatttg aagtgtctga 28020
aagtaaccaaa agatgtttac ttgagagtag ttatttctt ttagctctc agctctatac 28080
attcttccag ggaaccgtag atcttgggtc ctatttgagc cccaaaggat cagttagttt 28140
tacaaggac aatcgtattc tctgtcacat ctttttggc catgctcaa aagcagtcct 28200
acaatgtaag ctactgtca taggtcaat gcagtcacc ttcaaagcaa gagaataat 28260
ttcatgagta actccaactg ccgcttgtt atagggaagg catcatgttg gagcttcca 28320
gtcaaaatc tcacagtgaa caatttaagt ctaaagtca aaagtttcaa tggcatttgg 28380
tggaaaaaat atcactttac tgtgtacttc agacttcttg tactagtatt ttactatagt 28440
cagaagaaac atcattttt caagtatcac tttcttccc tctgtcttc aggaactgca 28500
ttgggcagga gtttgccatg attgagttaa aggttaacat tgccttgatt ctgctccact 28560
tcagagtgac tccagacccc accaggctc ttacttccc caaccattt atcccaagc 28620
ccaagaatgg gatgtattg caccigaaga aactctcga atgttagatc tcagggtaca 28680
atgattaaac gtacttgtt tticgaagt aaatttacag ctaatgatcc aagcagatag 28740
aaaggatca atgtatgtg gtaggtattg aggttgggtg gatagggtc tctgtgaaga 28800

10

20

gatccaaat catttctagg tacacagtgt gtcagctaga tctgtttcta tataactttg 28860
ggagattttc agatcttttc tgttaaactt tcactactat taatgctgta tacaccaata 28920
gactttcata ttttttctgt tgtttttaaa atagtittca gaattatgca agtaataagt 28980
gcatgtatgc tcactgtcaa aaattcccaa cactagaaaa tcatgtagaa taaaaatttt 29040
aaatctcact tcacttagcc gacattccat gccctgacca atcctactgc ttttcctaaa 29100
aacagaataa ttgggtgtgc attctttcag actttttcct atacatttta tatgtagaaa 29160
tgtagcaatg tatttgtata gatgtgatca ttccatattt gttattgatt tttttcactt 29220
aataaaaatt caccitattc cttatcattg ctttatggta ttctgtaata tgaatgtact 29280
ataatttatt taactatttt cttatttggg catttaagtt atttctagtt ttaaaaacat 29340
gcttgtcaat ggcaacaaaa gccaaaatg acaaatggga tctaattaaa ctaaagagct 29400
tctgcacagc aaaacaaact accatcacac tgaatgggca gcctacagaa tgggagaaaa 29460
tttttgcaac ctactcatct gacaaaggcc taatatccag aatctacaat gaactcaaac 29520
aaatgtacaa gaaaaaaca accccatcaa aaagtgggig aaggatatga acagacactt 29580
ctcaaaagaa gacatttacg cagccaaaag acacatgaaa aaatgcctat cgtcactggc 29640
catcagagaa atgcaaatca aaaccacaat gagataccat ctacaccag ttagaatggc 29700
aatcalttaa aagtcaggaa acaacagggt cggagagga tgggagaaa taggaagact 29760
tttacactgt tggggcagg agaatcactt gaacccggga ggggaggtt gcagttagcc 29820
gagggtggcg cactgcactc cagcctgggc gacagaacga gtactccatc tcaaaaaaaa 29880
aaaaaaagga caccaaactt ctcaatctta atgtgtcat ctatgtggta tcttcataa 29940
tcctctcag acagagtcct cttttgtga tatgatctta cagtattttt tgtttatacc 30000
attataatct cattaatgc agcaacacaa atgacaaaag acaactgatt tctcccttg 30060
gatgacctaa ttgtcttca ctcttccatc atcacttata acatgatgat tctcaaattc 30120
atctacctaa aatctatata taaaaaatc cctcccttga attccagatc ctggagaca 30180
aacaccacg tctaaaacca aatttgttta acactggacc agtcgtcctg tgtgactttc 30240
cattttgtca ctattttgtc agctgggtata ccaatatcca cccagttaa caatatitcc 30300
ttgttttttt ctggtacaaa cccaaataaa ttacaaacat caataaaagt aaaattctaa 30360
aataactcac tttctctata tatctcttc ttgtctgaaa aatgggttag gttagtcttt 30420
taaaagcatg catgataat tgtactgaat acaatattca ggtctggaca tactaggat 30480
aattttctgt gtctctggg tcttacctat ttgggicaa aataaacaag tttattaagc 30540

10

20

ttattaatat tcaatttcat tatcttcttt aacaattatg ticcctggta gtttcatlge 30600
 caataattta ttgtcaggt tgcaggigc ttctaaactt ctgtgtattt ttcatatcc 30660
 aattttactt taaatatatt tagaaaagag gtcgtttaa tttcctaata attattatat 30720
 tattgttttt tcactgacat ttgttgaatt gaaaaccctt aaaaatatga aatcattttt 30780
 tcaaaataig tgcacagac aattttgta aataagaaga cagaaacagg gcattatcaa 30840
 gagataaata ttcaatatac cttatatatt tgcacacat tttatacca acgttgccaa 30900
 aaattgtata tcatataaat gataacaagt tcacaaaggc attcctttat ccttaactc 30960
 tcaaattaga aactttcata ggtaggagt aggggaagca tatattccct ttgaaaggig 31020
 caagaaaatg tcattggcat tcaccatggt actcttcaag cttaaaaaaa atggactgca 31080
 aaacatttac aaacatagca tatttatgg gtacctttat gtttacataa atattgaaga 31140
 tatctacat accctttica atcagattat ctactgaca tttattgacc actttctatg 31200
 gggaaaac 31208

10

<210> 4

<211> 489

<212> PRT

<213> Homo sapiens

<400> 4

Val Ala Ala Leu Leu Gly Leu Leu Leu Leu Leu Lys Ala Ala Gln
 1 5 10 15
 Leu Tyr Leu His Arg Gln Trp Leu Leu Arg Ala Leu Gln Gln Phe Pro
 20 25 30
 Cys Pro Pro Phe His Trp Leu Leu Gly His Ser Arg Glu Phe Gln Asn
 35 40 45
 Asp Gln Glu Leu Glu Arg Ile Gln Lys Trp Val Glu Lys Phe Pro Gly
 50 55 60
 Ala Cys Pro Trp Trp Leu Ser Gly Asn Lys Ala Arg Leu Leu Val Tyr
 65 70 75 80

20

Asp Pro Asp Tyr Leu Lys Val Ile Leu Gly Arg Ser Asp Pro Lys Ala
85 90 95
Pro Arg Asn Tyr Lys Leu Met Thr Pro Trp Ile Gly Tyr Gly Leu Leu
100 105 110
Leu Leu Asp Gly Gln Thr Trp Phe Gln His Arg Arg Met Leu Thr Pro
115 120 125
Ala Phe His Tyr Asp Ile Leu Lys Pro Tyr Val Gly Leu Met Val Asp
130 135 140
Ser Val Gln Ile Met Leu Asp Arg Trp Glu Gln Leu Ile Ser Gln Asp
145 150 155 160
Ser Ser Leu Glu Ile Phe Gln His Val Ser Leu Met Thr Leu Asp Thr
165 170 175
Ile Met Lys Cys Ala Phe Ser Tyr Gln Gly Ser Val Gln Leu Asp Arg
180 185 190
Asn Ser His Ser Tyr Ile Gln Ala Ile Asn Asp Leu Asn Asn Leu Val
195 200 205
Phe Tyr Arg Ala Arg Asn Val Phe His Gln Ser Asp Phe Leu Tyr Arg
210 215 220
Leu Ser Pro Glu Gly Arg Leu Phe His Arg Ala Cys Gln Leu Ala His
225 230 235 240
Glu His Thr Asp Arg Val Ile Gln Gln Arg Lys Ala Gln Leu Gln Gln
245 250 255
Glu Gly Glu Leu Glu Lys Val Arg Arg Lys Arg Arg Leu Asp Phe Leu
260 265 270
Asp Val Leu Leu Phe Ala Lys Met Glu Asn Gly Ser Ser Leu Ser Asp
275 280 285
Gln Asp Leu Arg Ala Glu Val Asp Thr Phe Met Phe Glu Gly His Asp
290 295 300
Thr Thr Ala Ser Gly Val Ser Trp Ile Phe Tyr Ala Leu Ala Thr His

10

20

305 310 315 320
 Pro Glu His Gln His Arg Cys Arg Glu Glu Ile Gln Gly Leu Leu Gly
 325 330 335
 Asp Gly Ala Ser Ile Thr Trp Glu His Leu Asp Gln Met Pro Tyr Thr
 340 345 350
 Thr Met Cys Ile Lys Glu Ala Leu Arg Leu Tyr Pro Pro Val Pro Ser
 355 360 365
 Val Thr Arg Gln Leu Ser Lys Pro Val Thr Phe Pro Asp Gly Arg Ser
 370 375 380
 Leu Pro Lys Gly Val Ile Leu Phe Leu Ser Ile Tyr Gly Leu His Tyr
 385 390 395 400
 Asn Pro Lys Val Trp Gln Asn Pro Glu Val Phe Asp Pro Phe Arg Phe
 405 410 415
 Ala Pro Asp Ser Ala Tyr His Ser His Ala Phe Leu Pro Phe Ser Gly
 420 425 430
 Gly Ala Arg Asn Cys Ile Gly Lys Gln Phe Ala Met Arg Glu Leu Lys
 435 440 445
 Val Ala Val Ala Leu Thr Leu Leu Arg Phe Glu Leu Leu Pro Asp Pro
 450 455 460
 Thr Arg Val Pro Ile Pro Ile Ala Arg Val Val Leu Lys Ser Lys Asn
 465 470 475 480
 Gly Ile His Leu Arg Leu Arg Lys Leu
 485

10

20

【図面の簡単な説明】

【図 1】

図 1 は、本発明の薬剤一代謝酵素タンパク質をコード化する c D N A 分子又は転写配列のヌクレオチド配列を示す (S E Q I D N O . 1) 。さらに、ここには A T G 開始、終止、及び組織分布のような構造及び機能情報が示され、この分子配列に基づいた発明の特定用途を容易に決定することができる。図 1 の実験データによると、ヒトの胃、脳（乳児を含む）、子宮内膜腫瘍、前立腺、腎臓、副腎腫瘍、頭部／頸部、交感神経幹、胸部、及び肝細胞腫瘍で発現することが示されている。

30

【図 2】

図 2 は、本発明の薬剤一代謝酵素タンパク質の予測アミノ酸配列を示す (S E Q I D N O . 2) 。さらに、ここにはタンパク質ファミリー、機能、変更部位のような構造及び機能情報が示され、この分子配列に基づいた発明の特定用途を容易に決定することができる。

【図 3】

図 3 は、本発明の薬剤一代謝酵素タンパク質をコード化している遺伝子領域のゲノム配列を示す (S E Q I D N O . 3) 。さらに、ここにはイントロン／エクソン構造、プロモーター位置のような構造及び機能情報が示され、この分子配列に基づいた発明の特定用途を容易に決定することができる。図 3 に示されるように、S N P は 4 5 の異なるヌクレオチド位置で確認された。

40

【 1 】

```
1  GGGGCTGCG TGTCTGCG AGGCTGAG TGGGCTGCG ACTGCTTTC
61  CTCTTTCGCG GAGTACAGG GTTGGGCG GGGGCGAGG GGGGCTGCG
101  TGGGCGAGCG TGGGCGAGCG GGGGCGAGG GAGAGCGAG CTTCTGCGCG
151  GGGGCGAGG AGGCGGCGCG GTTGGGCGCG GAGAGCGAG TGGGCTTTC
201  CTGGGCGAG AGGCGGCGCG GGGGCGCGGT TACTGCGCG GTTCTGCTCT
251  GGGGCGCGGT GGGGCTGCG GAGGCGAGT AGGCTGAGT GGGGCGAG
301  GGGGCTGCG GGGGCTGCG GGGGCTGCG GGGGCGCGG GGGGCTGCT
351  GGGGCGCGG GAGAGTTCG TGGGCGAGT TGGGCGAGT AGGCTGAGG
401  AAGTATTCG AAGTATTCG GGGGCTGCG GTTCTGCGGT TGGGCGCGGT
451  GAGGCTGCTT TGTGATCTA TGGGCGAGT TGGGCGAGT GAGTCTGAG
501  GAGAGAGAT GGGGCTGCG GGGGCTGCG GAGTCTGCG GGGGCTGCT
551  TGGGCGAGG AGGCGGCGGT TGGGCGAGT GGGGCTGCT GGGGCTGCT
601  GGGGCTGCT GGGGCTGCT GGGGCTGCT GGGGCTGCT GGGGCTGCT
651  GGGGCTGCT GGGGCTGCT GGGGCTGCT GGGGCTGCT GGGGCTGCT
701  GGGGCTGCT GGGGCTGCT GGGGCTGCT GGGGCTGCT GGGGCTGCT
751  TGTCTGCGT TGTCTGCGT TGTCTGCGT TGTCTGCGT TGTCTGCGT
801  GAGAGAGAG AGGCGGCGGT TGTCTGCGT TGTCTGCGT TGTCTGCGT
851  AAGTATTCG TGGGCGAGT TGGGCGAGT TGTCTGCGT TGTCTGCGT
901  TGTCTGCGT GGGGCTGCG GGGGCTGCG GGGGCTGCT GGGGCTGCT
951  GAGTCTGAG AGGCGGCGGT TGTCTGCGT TGTCTGCGT TGTCTGCGT
1001  GGGGCTGCG GGGGCTGCG GGGGCTGCG GGGGCTGCG GGGGCTGCG
1051  GAGTCTGCG TGTCTGCGT GGGGCTGCG GGGGCTGCG TGTCTGCGT
1101  TGTCTGCGT TGTCTGCGT GGGGCTGCG GGGGCTGCG TGTCTGCGT
1151  TGGGCGAGG GAGTCTGCG GGGGCTGCG GGGGCTGCG GGGGCTGCG
1201  GAGTCTGCG GAGTCTGCG GGGGCTGCG GGGGCTGCG GGGGCTGCG
1251  TGTCTGCGT TGGGCGAGG TGGGCGAGT TGGGCGAGT TGGGCGAGT
1301  TGGGCGAGG GGGGCTGCG GGGGCTGCG GGGGCTGCG TGGGCGAGT
1351  GGGGCGAGG GAGTCTGCT GGGGCTGCG TGGGCTGCG GGGGCTGCT
1401  GGGGCTGCT GAGTATTCG GGGGCTGCG GGGGCTGCG GGGGCTGCT
```

```
1451  GGGGAGAGG AAGGCTGCT GGGGCTGCG GGGGCTGCG GGGGCTGCT
1501  GAGTCTGCG AGGCGGCGGT GGGGCTGCG GGGGCTGCG GGGGCTGCG
1551  GGGGCTGCG GGGGCTGCG GGGGCTGCG GGGGCTGCG GGGGCTGCG
1601  TGGGCTGCG GGGGCTGCG GGGGCTGCG GGGGCTGCG GGGGCTGCG
1651  TGGGCTGCG GGGGCTGCG GGGGCTGCG GGGGCTGCG GGGGCTGCG
1701  GAGGAGAGG GGGGCTGCG GGGGCTGCG GGGGCTGCG GGGGCTGCG
1751  TGGGCTGCG GGGGCTGCG GGGGCTGCG GGGGCTGCG GGGGCTGCG
1801  GAGTCTGCG GGGGCTGCG GGGGCTGCG GGGGCTGCG GGGGCTGCG
1851  GAGGAGAGG AAGGCTGCT GGGGCTGCG GGGGCTGCG GGGGCTGCG
1901  TGGGCTGCG GGGGCTGCG GGGGCTGCG GGGGCTGCG GGGGCTGCG
1951  GAGTCTGCG GGGGCTGCG GGGGCTGCG GGGGCTGCG GGGGCTGCG
2001  TGGGCTGCG GGGGCTGCG GGGGCTGCG GGGGCTGCG GGGGCTGCG
2051  GGGGCTGCG GGGGCTGCG GGGGCTGCG GGGGCTGCG GGGGCTGCG
2101  TGGGCTGCG GGGGCTGCG GGGGCTGCG GGGGCTGCG GGGGCTGCG
2151  GGGGCTGCG GGGGCTGCG GGGGCTGCG GGGGCTGCG GGGGCTGCG
2201  TGGGCTGCG GGGGCTGCG GGGGCTGCG GGGGCTGCG GGGGCTGCG
2251  GAGTCTGCG GGGGCTGCG GGGGCTGCG GGGGCTGCG GGGGCTGCG
2301  GGGGCTGCG GGGGCTGCG GGGGCTGCG GGGGCTGCG GGGGCTGCG
```

(SEQ ID NO:1)

FEATURES:

5' UTR: 1-189
Start Codon: 190
Stop Codon: 1720
3' UTR: 1723-2327

Homologous proteins:

Top 10 BLAST Hits

| | Score | E |
|--|-------|-------|
| gi 2117369 pir A29368 prostaglandin omega-hydroxylase (EC 1.14.14.1) | 521 | e-146 |
| gi 117166 sp P10611 CP44_RAB17 CYTOCHROME P450 4A4 (CYP1A4) (P... | 520 | e-146 |
| gi 164981 gb AAA31232.1 (J02818) cytochrome P-450-2 (Oryctola... | 520 | e-146 |
| gi 1656 emb CAA40493.1 (X57209) omega-hydroxylase cytochrome ... | 518 | e-146 |
| gi 89989 pir A34260 laurate omega-hydroxylase (EC 1.14.16.3) c... | 517 | e-145 |
| gi 117167 sp P14579 CP45_RAB17 CYTOCHROME P450 4A5 PRECURSOR (C... | 516 | e-145 |
| gi 203787 gb AAA41038.1 (M57718) cytochrome P-450 1A1 [Rattus... | 510 | e-143 |
| gi 89992 pir B34160 cytochrome P450 4A7 - rabbit >gi 164985 gb... | 510 | e-143 |
| gi 3738263 dbj BAA33804.1 (AB018421) cytochrome P-450 [Mus mus... | 509 | e-143 |
| gi 8393238 ref NP_058695.1 cytochrome P450, subfamily IVB, pol... | 508 | e-143 |

BLAST to dbEST:

| | Score | E |
|--|-------|-------|
| gb AW812435 AW812435 CM1-ST0181-261099-026-a02 ST0181 Homo sapi... | 1092 | 0.0 |
| gb R56515 R56515 yg94d06.r1 Soares infant brain INIB Homo sapie... | 769 | 0.0 |
| gb AA337301 AA337301 EST42040 Endometrial tumor Homo sapiens cD... | 640 | 0.0 |
| gb AA652746 AA652746 ns65c09.s1 NCI_CGAP_Pr22 Homo sapiens cDNA... | 636 | e-180 |
| gb AA863360 AA863360 oh04f03.s1 NCI_CGAP_Kid3 Homo sapiens cDNA... | 599 | e-168 |
| gb AA319338 AA319338 EST21550 Adrenal gland tumor Homo sapiens ... | 555 | e-155 |
| gb BP355963 BP355963 CM1-HT0878-060900-398-b08 HT0878 Homo sapi... | 381 | e-103 |
| gb BP445825 BP445825 nae41d04.x1 Lupski_sympathetic_trunk Homo ... | 365 | 5e-98 |
| gb AA557324 AA557324 n181a02.s1 NCI_CGAP_Br2 Homo sapiens cDNA ... | 357 | 1e-95 |
| gb AV683266 AV683266 GKC Homo sapiens cDNA clone GKCDQ... | 323 | 2e-85 |
| gb AW264444 AW264444 xr03d03.x1 NCI_CGAP_Brn53 Homo sapiens cDN... | 242 | 5e-61 |

EXPRESSION INFORMATION FOR MODULATORY USE:

Library source:

Expression information from BLAST dbEST hits:

gb|AW812435|Stomach
gb|R56515| Soares infant brain INIB
gb|AA337301| Endometrial tumor
gb|AA652746| normal prostate
gb|AA863360| kidney
gb|AA319338| Adrenal gland tumor
gb|BP355963|head neck
gb|BP445825| Lupski_sympathetic_trunk
gb|AA557324| breast
gb|AV683266| hepatocellular carcinoma
gb|AW264444| brain

Expression information from PCR-based tissue screening panels:

Whole brain

【 2 】

1 MEPSYLETVK APTPLAVV CLALGLQAI ELYLQRLL EDLQFFA77
51 TETFLGSDP IQDQKELR EIEKYPAY PFIQFPAV PCYEDP7AK
101 TLLSKTPES ETLQKSPYL LGEGLAALDG KETQQRLL TQCFE7WIL
151 ATIEVMASV EDLQKELR CSTQDTSVEV YEHKMSLD IIMCLAPSE
201 TSDQNTSTED PFAKATFELS EIIPEHLYSL IYEDSIFEL SPQGYRPOIL
251 SHYLSQYTDI IQDQKELR AGVECDNTPS ELYQDPLDIY LSADDES6SS
301 PSDIPHESEV STYLLAGEEDT LAASISVLY CLALAPEDQK RKEETV6IL
351 GQSSITVDQ LGEKSTTTC IETQCLIPA VPSISKLSL PLTFPDGCTL
401 PAGITVYLSI VEHETPAAY VEPKVPDPL EFSQESDQK EPTATLPFSA
451 GSRNCIGQEV AMIELKVTIA LILHEKVTY DPTRPLTPFN EYILEK6NG
501 TLEKELSEC

FEATURES:

Functional domains and key regions:

(1) PDOC00001 PS00001 ASM_GLYCOSYLATION

N-glycosylation site

206-209 ESTB

(2) PDOC00004 PS00004 CAMP_PDSFBD_SITE

cAMP- and cGMP-dependent protein kinase phosphorylation site

Number of matches: 2

1 265-268 KKKL

2 505-508 KKL

(3) PDOC00005 PS00005 PKC_PDSFBD_SITE

Protein kinase C phosphorylation site

Number of matches: 4

1 159-161 SYK

2 278-280 TPK

3 292-294 SAK

4 374-376 TCR

(4) PDOC00006 PS00006 CK2_PDSFBD_SITE

Casein kinase II phosphorylation site

Number of matches: 9

1 4-7 SYL

2 104-107 SRTD

3 172-175 STDP

4 176-179 TSTB

5 207-210 STED

6 292-295 SAKD

7 300-303 SFSD

8 302-305 SDID

9 393-396 TTFD

(5) PDOC00008 PS00008 MYRISTYL

N-myristoylation site

Number of matches: 5

1 25-30 GLQAI

2 298-303 GSSPSD

3 353-358 GSSITB

4 451-456 GSRNCI

5 457-462 GQEFAM

(6) PDOC00081 PS00086 CYTOCHROME_P450

Cytochrome P450 cysteine heme-iron ligand signature

448-457 FSA6SNCIG

Membrane spanning, signal, and domains:

| Helix | Begin | End | Score | Certainty |
|-------|-------|-----|-------|-----------|
| 1 | 12 | 32 | 1.638 | Certain |
| 2 | 76 | 96 | 1.029 | Certain |
| 3 | 316 | 336 | 1.077 | Certain |
| 4 | 395 | 415 | 1.443 | Certain |

BLAST Alignment to Top Hit:

>gi|2117369|pir||A29368 prostaglandin omega-hydroxylase (EC 1.14.15.-)
cytochrome P450 4A4 - rabbit
Length = 510

Score = 521 bits (1328), Expect = e-146

Identities = 246/493 (49%), Positives = 355/493 (71%), Gaps = 1/493 (0%)

Frame = +1

Query: 235 LAFYFCLALGLLQAIKLYLRRQRLLRDLRPPFAPPTHWFLGHQKFIQDDN-MEKLEEIE 411
+A + L L LL+A +LYL RQ LLR L+ FP PP HW LGH + Q+D +E++++ +E
Sbjct: 21 YAALLGLLLLLKAAQLYLRQWLLRALQOPPCPPFWLLGHSREFQNDQELRIQKWE 80

Query: 412 KYPRAPFFWIGPFAFFCIYDPDYAKTLLSRDPKSRYLQKPSPLLGKGLAALDGPKW 591
K+P A P+W+ +A +YDPDY K +L R+DPK+ K P +G GL LDG WF
Sbjct: 81 KPPGACPPWLSGNKARLLVYDPDYKVLGRSDPKAPRNYKLMTPWIGVGLLLDGGTWF 140

Query: 592 QHRRLLTPGFHFHNLKAYIEVMAHSVKQDLKWEKICSTQDTSVEVYEHINMSLDIIMK 771
QHRR+LTP FH++LK Y+ +M SV++MLD+WE++ S QD+S+E+++H++ M+LD IMK
Sbjct: 141 QHRRMLTPAFHYDILKPYVGLWVDSVQIMLDRWEQLIS-QDSSLEIFQHVSLMTLDTIMK 199

Query: 772 CAFSKETNQTNSTHDPYAKATFELSIIIFHRLYSLLYHSDIIFKLSPQGYRPOKLSRVL 951
CAFS + + Q + Y +AI +L+ ++F+R ++ + SD +++LSP+G P + ++
Sbjct: 200 CAFSYQGSYQLDRNHSYIQAINDLNVLVYRARNVHQSDFLYRLSPGRLFHRAQQLA 259

Query: 952 NQYTDITIQERKKSQAGYKQDNTPKRKYQDFLDIVLSAKDES6SSFSIDIVHSEYSTFL 1131
+++TD +IQ+RK LQ + + +++ DFLD++L AK E+GSS SD D+ +EV TF+
Sbjct: 260 HEHTDRVIQQRKAQLQGELEKVRKRRLDFLDVLLFAKMENGSSLSQDLRAEVDTFM 319

Query: 1132 LAGHDTLAASISVILYCLALNPEHQRCREEVRGILGDCSSITWDQLGEMSYTMCIKET 1311
GHDT A+ +SWI Y LA +PEHQ RCREE++G+LGDG+SITW+ L +M YTTMCIKE
Sbjct: 320 FEGHDTTASGVWIFALATHPEHQRCREEIQCLLGDCASITWEHLQMPYTTMCIKEA 379

Query: 1312 CRLIPAVPSISRDLKPLTFPDGCTLPAGITVYLSIWGLHNPAAVWKNPKVFDPLRFSC 1491
RL P VPS++R LSKP+TFPDG +LP G+ + LSI+GLH+NP VW+NP+VFDP RF+
Sbjct: 380 LRLYPVPVSVTRQLSKPVTFFDGRSLPKGVILFLSIYGLHNP-KVWQNPVFDPRFAP 438

Query: 1492 ENSDQRHPYAYLPFSAGSRNCIGQEFAMIELKVTIALILLHFRVTPDPTPLTFPNIIFIL 1671
+++ H +A+LPFS G+RNCIG++FAM ELKV +AL LL F + PDPTR +L
Sbjct: 439 DSA--YHSHAFLPFSGGARNICIGQFAMRELKVAVALTLRFELLDPDTPYPIIARVYL 496

Query: 1672 KPNKGMVYHLKXL 1710

K KNC++L L+KL

Sbjct: 497 KSKNGIHLRLRL 609

Hmmer search results (Pfam):

| Model | Description | Score | E-value | M |
|----------|------------------------------|-------|----------|---|
| PDOC0067 | Cytochrome P450 | 416.5 | 2.5e-121 | 1 |
| CE00363 | EG0363 glycine_receptor_beta | 2.1 | 4.7 | 1 |

Parsed for domains:

| Model | Domain | seq-f | seq-t | hmm-f | hmm-t | score | E-value |
|----------|--------|-------|--------|-------|--------|-------|----------|
| CE00363 | 1/1 | 210 | 233 .. | 481 | 504 .. | 2.1 | 4.7 |
| PDOC0067 | 1/1 | 46 | 504 .. | 1 | 497 | 416.5 | 2.5e-121 |

【 3 】

1 CGAGCTCTC TTAGGCTCTT AATATAGT CAAAAGTTC CAGAGTTCT
 51 TTCTTADCA TGAAGGACA TGAAGGCTG CTGACAGGG GGAAGTGGC
 101 CTGAGGAGA GAGTAAGTC CATAGAACTG TCGAGGCTC AGAGGAGTC
 151 ACAGCAGAG CAGGAAGCTG GGTGGAGTA GGTGAGGAA GGGTTTCCA
 201 GGTCTGAGC CTGCAAGAG AACTCAITAG AGGTGACCA ACCACACATA
 251 CTATTCTCG GTCTCATGA GAGCCAGGG AGGAGCAGG GCAAGATATC
 301 ACAAGCTGA AGTTTCAGT CTGGGACGA GCATGGATCT CAGTCTTTG
 351 GGGTACGAG CATGGATCA TATGAGGGC ATCATACAG CATCATGAT
 401 TGGGAGGCA ATAGGAGTA GAGGAATCAT ATGAAGGCT GAAATGGAT
 451 GAGTTAGCA CAGAGAGTC TGTAGGAGG AGGATTTCTA GAGGCTGCA
 501 AATAACAGA TCTAGTTGA GGTTCAGTT AGTACAGAG TAGGAGAGT
 551 TGGGAAGAA GAGGTGAAA CAGTTGCTG GTGTGGCTTA ATGGAACATG
 601 CAGGGGCAA GAGCAAGTT GGTGAGATG AAGTACGAG GGGTGGGGC
 651 GTGTCTTTT TTTTCTTTT TTTTCTTTT TGAAGGAGG TCTGACTGTG
 701 TCAGGAGCT GAGTGCAGT GGGGAGATC GGGTCACTG CAATCTTTG
 751 CTGTGGGCT CAGGAGTTC TGTGCTCTA GGTCTCTGAG TAGTGGGAT
 801 TACAGGGGG GGGAGGAGG GGGAGCTAT TTTAGTACTG TTAGTAGAGA
 851 TGGGTTTCA GATCTTGGC CAGGATGCTG TTGATGCTT GAGTCTGTA
 901 TGGGGGGGG TGGGCTTGG AATTTGCTG GATTACAGG GTGAGGAGC
 951 GGGGGGGGG GGTGAGGAG TGTCTTAATG ACTTACGGC CAATATAAT
 1001 CTGGTGGAG AGAGTGGAG GTAGGCTTA GGAATTTGG GGGAGGGGG
 1051 GGGAGAGTG GGGAGGAGG AGTGTAGGG AGAGAGGGA ATTGTAGCAG
 1101 AAGTTGGGT TATTGTTAG AGGTGTCAAT GAGAGCTTA CATATGGCTG
 1151 TCTTAGGCTA AATCAATGA TAAATGATG AATAATATA TGAATGAAT
 1201 GTGGCAATG CCAATAGGA TTGCTGGAG AGGAGGCTG GGGAGAGAG
 1251 CAGCTTGGG AGTCAAGGCT GTTAGATCT AGTTACGAG CTGATAGCT
 1301 ACAATACTA AAGCAATCAG TTTCAAATTA TTTTACTAG ATTCTCTGT
 1351 TATCTGACT GAGTTTAT TATGTTCTG GCATCTAGG TGAAGGCTC
 1401 ATGGGAGTA GAGGAGGCA GAGGAGGAG GGTGTGAGC CAGAGAGCA

2901 TTGCTCTCT GGTCTTTTA AATCACTGC CAGGCTAAT TAGGTTTTT
 2951 ATTCAATCA GAGCTCTGA GTGGCAATG TGTGATAGT ACTAAGATTA
 3001 CAGAGAGAG CTAAGTGGT GGTCTGAGC AGGAGGCTA GTGAGAGCT
 3051 TAGGAGCAG AGAGAGAGT AAGATTTAG GCAATGGTG CTTTACTAG
 3101 GGTCTAGAG CAGAGGATA TGTGTGGAG GAAATATAC ATCTGGGCT
 3151 AGAGAGGAA GGAAGTCTG TGAAGGCTG AGGAGGCTT TAAAGGATG
 3201 TTGGGCTGT TGGGAGGAG ATTGAGGAG AGTACTACA GATCTCTTG
 3251 GTTGGGAGC TTCTAGTCT TTCTATATA AAGCAAGAG TTGAGCTCT
 3301 TTTATGGCT TCTTCTGTA TTTAATAGT TATTGTTAA ATAGTATTAG
 3351 CATATTGAT CTAATTAAT TAAAGTTTA GATATGCTT GTGTTTACA
 3401 TATGTTTGT TGTGCGGAG CAGGCTCAT GTTGAATTT GATTGCAAT
 3451 GTTGGAGTG GATGTGATG GAGATCTTT GGTCAATGG GATGATGCG
 3501 TCATGAATG CTGCTGAGC CTGCTGCTT CATAGTTCT CACTCTCTA
 3551 GGTCTCTTC AAGGAGGAG ACTGATGTT GAAAGAGGC TGGAGGCTC
 3601 TGGGCTCTT CTGCTGCTT CTGAGATGT GGTCTGCTA CAGAGTGT
 3651 GGTCTGCTT TCAATATGA TGGAGGAGG TGTGAGTGC TGGAGAGT
 3701 CAGATGCAA TGGATGCTT CTGAGAGG CTGAGAGAT GTAGGAGAA
 3751 TAACTGCTT TGTGATGAG CAGGCTCAG GTATTGCTT ACAGGAGAG
 3801 AATGTACTA AGAGAGATC CAGTATGAA CTCTTTATG AGAGGAGATC
 3851 ACTTACACT CATATTCAG TGTGAGGTA ACTATATAG ATTGTATTT
 3901 TTAATAGAA AAGCTTCTAT TTGATTAAT TTTATATGC AATGTTTAT
 3951 TACTGCTAT CTAATGCTG CTCTTTTAT TTTATGCTT TTCTATAGA
 4001 ACTTTTGGC CAGGAGGCA GTATTGAGG GGTGAGGAG GGTGAGGAG
 4051 GGTGAGGAG GGTGAGGAG GGTGAGGAG GGTGAGGAG GGTGAGGAG
 4101 GGTGAGGAG GGTGAGGAG GGTGAGGAG GGTGAGGAG GGTGAGGAG
 4151 GGTGAGGAG GGTGAGGAG GGTGAGGAG GGTGAGGAG GGTGAGGAG
 4201 GGTGAGGAG GGTGAGGAG GGTGAGGAG GGTGAGGAG GGTGAGGAG
 4251 GGTGAGGAG GGTGAGGAG GGTGAGGAG GGTGAGGAG GGTGAGGAG
 4301 TCAATGAGC AATAATGTA AGGTCTATT CTGAGTCTT AATCTAATC

1451 TATGCTGGT TTAATGCTT GTATCTTAT AATTTTAA TAAATTTTA
 1501 TGGGCAAT TCAATTTGA CTGAGATTA TAAATTAAT AGAGGCTCT
 1551 GAGTCTAGT GTATAGTGA ATTCTATAT GATGCTAGC TACTGTGAT
 1601 ATCTTGGGG TTGAGTCTT AGTGGCTCG TATGGGAGG TTGAGTAGC
 1651 TATGCTACA GGTGGGAT CAGGCTGGA ATCAATTTA AATCAATTTAC
 1701 GGAAGAGCA CTAGGAGGC CAGAGATTA AGGAATATG ATGCTAGAGC
 1751 TGGGCTAGC TGTAGGCTT GGAATTTTG GTAGATGCT AGAATGAAA
 1801 AGAATATCT TTGCTAGGC ATTTATAT TGTGATAGG AAGAAAACA
 1851 ATGAGCTAG CTTTAGCAT ATTTAGCAT ATAAATTCAG ATGCTAGAGC
 1901 TGAATATCT TGAATTAAT AGAGGAGCT CAGAGGCTC AAGAGGCTT
 1951 GGGGAGGAG AGGCTGGG GGTGCTTA CAGGCTAGG ATGCTAGAGC
 2001 GGGGCTGGC TGTCTGGC AGGCTGAGG TGGGCTGGC AGTGGCTTT
 2051 GTTCTGGC GAGTCAAGG GGTGCTAGG GGGGAGGAG GGGGCTGGG
 2101 GTGGGAGC CTAGGAGG TGGGAGGAG AGAAGGAGC GGTCTGGC
 2151 GGGGAGGAG AAGGAGGAG GGTGCTAGG GGGGAGGAG ATGAGATTT
 2201 CTGCTGAG CAGGCTGG GGGGAGGAG GTTCTGGC GTTCTGGC
 2251 TGTGCTGG TGGGCTGG CAGGCTAGG AGTCTAGC TGGGAGGAG
 2301 GGGGCTGG GGGGCTGG GGGGCTGG AGGCTGGC AGGCTGGC
 2351 TGTGCTGG CAGAGGTA AATGAGGAG AAGAGGTA AAGAGGAG
 2401 AGAGGAGG CAGAGGAG TGGGAGGAG GGGGAGGAG GGGGAGGAG
 2451 GGGGCTTT TGTAGGCTT GGGGAGGAG GGGGCTGG AGGCTGGC
 2501 AGGCTGGC GGTGCTTA GATCAATTT TGTGCTGT GGGGATTT
 2551 TTTGGGAG GGGGAGGAG AAGGAGGAG AAGAGGAG GTTGGGAG
 2601 TGGGAGGAG CATTTAAG AAGGAGTA TGAAGGAG AGGAGGAG
 2651 TAACTAGG CTGCTGTA TGTGCTAG AGGCTTGA CATGCTGG
 2701 TTAAGGAG CAGTCTAT CTAATATCT TATTAGTTC AGAATATG
 2751 CAAATCTAT CCAATGCTT GGGGCTTTA GTGCTTTA AGAGAGAA
 2801 ATTTCTTTT CTAAGATTC TGTGAGTA AGGAGGAG CAGGAGGAG
 2851 ACTGAGAA TACTGATTA TGAATTTG TAAAGGAG CAGTCTGG

4351 CATTGATTA TATGCTAT CTAAGTAT GAGAGAGAG GTAGAGGAT
 4401 GGTAGGAG GGTGGAGG GATAGAGGAG AGTGGGAG GAGGAGGAG
 4451 AAGGTTAAT GGTAGGAG AATAGAGAG AATGATTA ACTCATTT
 4501 TGTAGGATA GAGGAGGAG TATAGGAT AATAAGTGA CACTTTTAA
 4551 TAAAGGAT TATAGGAT TTGCAATCT AATGATTA TGTGAGGAG
 4601 GATGAGTGC CATTCTTA TGTGAGTCT ATTTGATTT GATGAGTGT
 4651 ATCAAGGAG TGTGATTA TGTGATTA TATAGGATA CTATGATTC
 4701 AGAGGATTA AATGATTA ATAAATTA TATAGGATA TGTGATGT
 4751 AGAGGAGC TGTGATTA TGTGATTA TATAGGATA TGTGATGT
 4801 GATGAGTGC GGTGATTT GGTATTTTC AAGATTTAT TGTGATTT
 4851 GATGATTA TGTGATTA AATTTAGAG TGTGATTA AATTTAGAG
 4901 AGAGGAGG CAGGAGTCT GGTGAGGAT GATGATTA TGTGATTA
 4951 TTTGGGAT TGTGATTA TAAAGATAT AGGCTTTCT ATGATGAG
 5001 AGAGGAGG TTTGGGAT GTAGGATAT CTTAATTTT TTTGATTT
 5051 TTTTGTGT TTTGAGAG GATGATTT CTGAGGAG GGTGAGGAG
 5101 GATGAGGAG ATGAGGAG AGTCAATCT GGTGATTA GATGAGGAG
 5151 ATGAGGAG CTAGGAGG CAGGAGGAG GATGAGGAG CAGGAGGAG
 5201 CAGGAGGAG AATTTTGA TTTGAGGAG AGAGGAGG TGTGATTT
 5251 GGTGAGGAG GGTGAGGAG GGTGAGGAG TGTGAGGAG GGTGAGGAG
 5301 GGTGAGGAG TGTGAGGAG GGTGAGGAG GGTGAGGAG GGTGAGGAG
 5351 ATTTTGTTA AGATGATTT TGTATTTTC AAGATATTA TGTGATTT
 5401 TTTTGTTA ATTTATTT TTTGATTT TTAATTTAT TGTGATTT
 5451 TTAATGAT CATAGTGT TAGAGTCA ATGCTCTTT GATGATTT
 5501 AAGTTTGT TTTTGTGT TGTAGGAG TGTATGAG TTTGAGAT
 5551 CAGAGGAG GATGATTA AGTGTGTA TATTAGGAG GGTGAGGAG
 5601 TATGAGGAG GATGATTT GGTGATTT CAGGAGGAG AATGATTT
 5651 CAGGAGGAG AAGGATTA GTTGTGTA TGTGAGGAG CATTTATTT
 5701 CTAGATTA ACTAGATTT TATGAGGAG CTGATATTT TGTGATTT
 5751 CATAGATTT GATGATTA AATGAGGAG TTTAGGAT TGTGATTT

5801 TAACATTTAT TATTTATTTG TGTTCGAACT ATTTCAAGTC TCTTCAGCT
5851 CTTTCAGAAAT ATTTCAATACA TTATTTGTAA CAGTCTATTT CAACACTGCA
5901 ACTTATTTCT TCTATCTAAA GACAGTAAGA TTTTAAATAT AGTCAATAGG
5951 TTACAGAGGG ATAAAGTGTG TATAGGCAAA ATTCCTTACA AGATGAGAAAT
6001 TTCAATTCCTT ACTTCTAGTA ATACAGGTCT TCAACATGCG CAAGGATATT
6051 CTTCTCTGCG AGCTTTGAGC ATCCAGGTCT GTGCTTATAT TCTCTCTCT
6101 CCAATTTAT TCTAAAGAGC CTTCTCTGCG ACCTTTGAGA CTAAATAGG
6151 ACCTCTAGTA CTTCTATCTT CCAACCTAT TATTTATATC TTGCTCTCTA
6201 TCACTCTGCG ACCTATAGCT CTATCTCTT TTCTCTCTCT GTTATTTATC
6251 CACCACTAAC TACAATATAA AATCTGTGAG AGTATAGATC TTTCTTTGCG
6301 ACTATAAAGC TACTGCAATG TACAGTTCTT CTTCTCTAT TACTCTCTCA
6351 TAAATCTCTT GTTGAATGCA TAAATATATT AGTCTGTGAG AAAATTTAT
6401 TATTCAGAGA TCAATTTACT GCATAGATA GCGCAGGTCT TTTGCAATTT
6451 ATTCATAGC CACATATAGC GACCTAGGAT GTACATATGC AGTCTGTCT
6501 GTCTATGCTG GTCTGATCTT GCATGTGTAC TTGATGTAC TCCAGAGAAC
6551 ATCTATGAG CTAAATAGTA TAAAGCACTT GCGCTCTGAG GTTAACTGCG
6601 AGTTTGAATC CTAATAGTG GTTCTGAGCT GTACACACTT GCGCAGATCA
6651 TTTAAGCTAG TCTGTAGGCG TCAATTTCTT CATCTCTAAA GTAGGATTTG
6701 TAAATCATAT TACTTCATAG GCTTCTGAT GTAAATATTA AATAACATAG
6751 AACATGAAA GCATTTAGCA GCACCTAGTT CATAGAGTCT CTTGATTAAT
6801 GTTCTCTGCT GTATTTGCG GCGCTATGCG ATTTCTGAA CATTTCTGAA
6851 CAATGTTTAC TAAATATATG TACTAGCTCT TTTCAAGTCT ATTTAGATGC
6901 TTTCTCTGCG ATGAGAGAAAT ATAAATTAAT TATAGTACAG TATTCACAGC
6951 AGTTTCTCTT CTTTCTCTG TACTAGGAG TTACAAAAG TATAATGAAA
7001 TACTTTCTAT TCTCTGCTCT GTTTATGAAA ATTTTCTACC TAAACAGACA
7051 ATTTCTATAT TACTTTACAA TATTCATGAG GCGAAGGCG TTTCTCTCT
7101 TATATTTCTC TCTATCTCTA CCACTCTGTA CTTCTGATAG ACAATAAATA
7151 CTTCTCTCTT TATTTCTCTT AAATGATTA ATGAAAAAT ATTCACATTT
7201 TTTAAAAACA CTACTCTGCA TACTAGTCTG GTCTTATCA CTGCTTCTAT

7251 TATGCAACA TTAACAAAA AGTCACTAT TTTAGAACT AGGTTTCAAG
7301 ACTCTCAAC TTTCACTGCG CTTCACTAT CAGAGAGACA CTTATGCTG
7351 TAAATTTCTT AAATGATAG AGAGAAAAAT GCGAGGAGA GTTCTGCTG
7401 TCTCTAGAGG ATGAGAGCAA ACAATCTGCG AGCAGATAGC GTTCTGCTG
7451 TCAAGAGAGG AAAATCTGCG CAGCTTAAAG ATTACTTTCT AGCGGCTGAG
7501 TCCAGAGATC ATCTCAAGT GTCTAGATG CAGACAGAAA AGACAGACAT
7551 CAGGCTGCG TCACTTTAC ACAAGAGCA AATGATTAAT AATCTTTCT
7601 TCACTGATTT AGAGAAAAA TTTAGTCAAG CCAATTAACA TCTCTCTTA
7651 AAGCAAAAA CTTCTCTCTT CTTCTGATTA TTTCACTCTT CATTTCTAA
7701 TTTCTGAGG TCAAAATTTT GTATTTGAT TTTCTTAAAG ATTCAGACAT
7751 CTTCTGAAAT CCACTTTCTG TTTCTGCGCG GAGATTTCTG ATTTCTTTA
7801 TACTAGAGT TCGCAATTTT TTAGAGAGAA GCATTTACTG CTAAGTCTG
7851 AGAATTAATC ACTGCTGAT AAATAGAGAG AGCAAGAGA AGAAGAAATG
7901 GTGAGCTGCA TCTAGGCTCA TCGGCAATTT AGTACTGTT AGTTCTGAG
7951 ATAGAGAAAT CTTCTTTTAA CTTCTGAGAT CCACTCTCAA TCTGAGTCTG
8001 TCACTTTCTG AGTCTGCGCA GAGAGTCTGA CTTCTCTGAG CTTCTATTTG
8051 GAGAGAGTT CAGCTAAAT GTCAATTAAC AGTCTCTGCG TCTGAGGATA
8101 CAGGATATAT TAAATTTGAG AAGATAAATA CACTAACTTT GTTAGAGAA
8151 TTAATCTCTA AGCTTGAAG TCGCAAGAA AGCATTTAT GTCACTTCA
8201 GAAAGTCTC AGCTCTCTT CTTCTGCTGA CTTATCTAG TCTGAACTC
8251 AGCTTCTGCT TATAAGAGCG CAGAGCTGCA GTTCTGCTG CTAATTAAT
8301 TTAATTTTCT CACTGAGTT TATTCAGAT GATAAGATG AGAGCTTCA
8351 GCAATTTAT GAAATTAAC CTTCTGCTCT CTTCTTCTG ATTTGCTCT
8401 TTTGAGCAT TTTCTGATC TATGAGGAG ACTATGAAA CAGCTCTGCG
8451 AGCAGAGAG GTAGAGAGAG GCGCAAGCT CTTGAGCTTA TTTCTCTGAG
8501 AAGTGAATG CATAAAGGCG ATAGGAGAGA TTTCAAGAGA AAGATTTCT
8551 TCGGCTCTT AGAGAGACA CAGAGAGTA TCGGAGGCT AGAGTTTCT
8601 TACCAATCT GAGGAGAT CCAATCTCT CCGAGTCTG TTTCTTCTG
8651 CAGATATCA TCGGAGGCT GAGTCTGTA TAACTTAAAG CTAAGCTGCG

8701 ATTAGCAGAC TTGCGAGGCA AGCTTCTCT TCGCTCTCT GGGTTTATG
8751 ACTTCAAGT CAGCAAGCT TCGCTCTCT AGCTCTCTG TCGAGCAATA
8801 GTCTCAGAG GGTGAGAAAT CAGCAGTAA TCTACTCTG CTTCTCTGCT
8851 ATGAAAGCTT GAATGCTAGA TCAATTAAT ACCATCAGA CTTCTCTGCT
8901 TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT
8951 TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT
9001 TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT
9051 TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT
9101 TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT
9151 TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT
9201 TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT
9251 TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT
9301 TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT
9351 TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT
9401 TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT
9451 TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT
9501 TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT
9551 TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT
9601 TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT
9651 TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT
9701 TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT
9751 TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT
9801 TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT
9851 TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT
9901 TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT
9951 TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT
10001 TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT
10051 TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT
10101 TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT TTTCTCTCTT

10151 TTAGGTCAG ATTTCAAGAG CAGCTGCGCG AACATGCTCA AATCCATCT
10201 CTACTAAAA TACAAAAAT AGCGAGGCTT GGTAGATCT AGGCTCTGAG
10251 TCGCTACTT TCGAGGCTG AGCAGTGAAG ATCTGTTGAA CTAAGAGGCT
10301 GAGGCTGCT GTAGCTGAG ATTTCTGCTG TCGCTCTGAG ACTAGGTCAG
10351 AGAGCAAGC TCTCTGCGCA GCGGCAACA ACAAAAAA AACTACCAA
10401 ACTGAGTCT CAGCTCTCTT ATTTCTGTT TTTCTCTCTT TTTCTCTCTT
10451 TTTCTGATG TTTCTCTCTT TTTCTGAGT CTTTATTTT CAGCTGCGAG
10501 TCAATAAAT TTTCTCTAG AGTTCTGCAA TATTTCTGCA CAGATAGAG
10551 TAACTCTTT ATTTCAATTA TTTCTATTA TTTTAAAGAA CAGATAGAG
10601 AAGAAAAAG ATATAGGCTT GTATTAGCA CCACTTAAAT TTTCTCTG
10651 CCGCTCTCT CATTGCTAG TTAATATGA TTTCTGCGAA GTCTTTAT
10701 CTTCTCTAT CTTCTCTCTT TCACTTTTGA CAATAGGAT CAGCTCTGAG
10751 TTTCTGCGCT GTTATAGGA TTAATGAAA TACATATTT TACCTCTGAG
10801 TTTATGCGC ACCATTTAT GAGTCTGAG TCACTCTGCA TCACTCTGAG
10851 ATCTCTGAG CATTCAATGA ATCTATCAA TTTCTATTA ATAACTTTA
10901 AAGCTCTTT AAGGCAAAAT CTTATGAGC AGTCTGCGAA TACAGATTT
10951 TTAAGCAAT CTTCTGCTAA GTATTTCTG ATAGTATCAT TCACTGAGCT
11001 GCGCTCAAG AGCAATATC AGCGAGAGA GTGAGACTC TTTCTCTCTT
11051 CTTCTGAT TTTCTCTCTT CATATAGTT ATCTATGCT TTTCTCTCTT
11101 AAGCAATTA CTTCTCTCTT TACTAGTCT AGCTAGGAG TACTGCTCT
11151 ACCTATGTA TTTCTCTCTT GTTAACTCT AGAGCAAGC AGGCAAAAG
11201 GTATTTAGTA TCACTCTCTT TTTCTGAGAA AAGTAAAGCT CAGGAGATTT
11251 CAGTCTCTTA TTTCTCTCTT ACATAGTAG TAACTCTCT TTTCTGAGTT
11301 AGCTCTCTTA TTTCTGAGG CTTCTGAAAT TATTTAGCT TTAAGCTCTT
11351 TTTCTGAGTA TTAATATAT CCACTCTGAG GTTCTCTCTT TAAATCTCT
11401 TTTCTCTAT AATTTCTCTT CCACTATAT AGCTATCTA CTTCTGATCT
11451 CCACTCTCTT ATCTCTCTT GTGCTCTCT GTTCTCTCT CTTCTCTCTT
11501 TTTCTCTCTT ATCTCTCTT TCACTATAT AATGAGATA AGGTTCTCTT
11551 CTTCTCTCTT CTTCTGAGAA AGCTATGAG CTTCTGAGC AGGCAAAAG

11601 GTTCCAGCAT GTGCGCTAC TAACCTCTG ATTCCATTTT AACATCTGA
11651 AGCATATCAT TGAGTGTATG GCTCATTCTG TGAATAATAT GCTGTAAAT
11701 AAGGCGGCAA AGTGTCTGTG GCAATTCGAA ATGCTCCAG CAATGACAG
11751 TATTAGATAT GTGTTTCTG GGCATGAA ATAAAAATC AGTTCTAAA
11801 AATTAAACA ATGTACAGT ACTTATGAA CAATAGCTGT CTGTAAAAA
11851 TTTGTTATGT TCTTGAAGT ATAATATTA TAAAAAGAT TGTGCTCTG
11901 TCTTACATAT ATTTTGAGT TTTATGCGAG CAAGCAAGT ACCAAATGT
11951 CATAGTTAGA TAGTAAGTGT TGTAGATCTG TTTTATGAG GCGGCTCTG
12001 TACAAACCTA CCGCAAGCT TGAGCAAGT GAGAGCTGA ACAAAGAGC
12051 TCAAGTTTC TTAAGAGAA ACATTCATA GAGCTTTCA AACAAAAAC
12101 ATTGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG
12151 AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG
12201 AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG
12251 AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG
12301 AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG
12351 AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG
12401 AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG
12451 AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG
12501 AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG
12551 AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG
12601 AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG
12651 AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG
12701 AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG
12751 AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG
12801 AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG
12851 AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG
12901 AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG
12951 AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG
13001 AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG

13051 AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG
13101 AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG
13151 AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG
13201 AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG
13251 AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG
13301 AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG
13351 AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG
13401 AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG
13451 AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG
13501 AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG
13551 AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG
13601 AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG
13651 AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG
13701 AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG AGCGCGGCGG
13751 AGCGCGGCGG AGCGCGGCGG GTGAGGCTTT GCTGCGGCGA GCTGCTGCA
13801 AGAGCTGCTC TGCACACTG CTGCTTTCT CACTTTTGA TCCAAAGCT
13851 TTTGAAATG TGTGAGTTT ATTTAAAT GTGCTATGG TGTGAGAG
13901 CAGTGCAGG GTAGTAGCA AGTTTGAAT TGAATTTGA GGAAGGCTG
13951 GGGTAAAGC CTGTAAATG TGGTCTGTG GTCAATGAT GCTTTAATG
14001 AGCTCTGCT TGTGAAAG CAGATTTAT GTAAATATG AAAAGGCGA
14051 GATCTTTAG TCAAGCAAT ACCATATAT CAGTTTCTC CATGCTCTT
14101 CTGCTGCGG TGGTGTATG TTTGCTGCT TGTGCGCTG TGTAAAGCA
14151 TGGTATATT ACTCATGTA TCTTGTCTC CTGCTGCTC AGGCTTCTT
14201 CCAATAGAT ATAAAGAGG GCGAGGCGG GAGCTTTCA ATGAAGCAA
14251 TTTGCTCAT GTGCTGCTA TCAATGCTT CTGAGCTCC TGTGCGAGG
14301 TAAGTGGAG AAGATTTGA GCACTGAGA CAGAGGCTG GAGTCTATG
14351 AGCATATCA CTGATGTCT CTGATATAT TCAATAAAT GCTTTTACG
14401 AAGAGAGCA ACTGCGAGC AAACAGCTA GTGCTGGAG AGCAAAAG
14451 ATATTCTCT ACATTTCTA AGTTCTTAT TAACAGATTA TCCAACTTT

14501 CTCTTCTAG ACCATGATC CTATGCAA AGCATATTT GAATCAGCA
14551 AATCATATT TCAGGCTTG TACAGTTGT TGTATCAGG TGACATAAT
14601 TTCAACTCA GCGCTCAGG CTACGCTTC CAGAAATTA GCGAGGTTT
14651 CAATCAGTAC ACAGATATT GTTGGTTTG GTTGGCCAC GTCCATAGC
14701 TGCCATGAT GTACTGTCT TGTCTAGAG GATAAGCTT AATATGCAA
14751 CAGAAAGAT CTTTGTTAT AATGAGCTT TTATATAGC ACTGCTCAA
14801 AGAAATTTA CTGAGTCTT TTATAGACT TTGCTTCAAC CATAGAGTA
14851 TTATCAGAT TTTTATAT ATATATATC ACTATTTTA TTATGAGCA
14901 TTATATTAA TACAATATA ACTAGGACT TANGAGTCC AGCATATAT
14951 GCAATATGC TTTTTCACA GCGATTCAG TAATAATAT CACAAGCTA
15001 AAGCATTCAT GCAATAGC AATGAGAGT GCAACAGAT AACATGAG
15051 ATGCAAGCA AAGATATTG ATTCAAAAC AGTTTATGA AGCATAAAG
15101 CAAAAGTTA AATAGATTA GCTTTTAA GAAATCAGA TTACTTGTA
15151 TGAATGCAAT AATGAGAGT ACTTATCAG CAGTGAATTA ATGCTTCTC
15201 AGCTTACCA CTACTAGCA GTTACTAGT AGTTACTTAC TGTCTTCTT
15251 CAGTGTCTC TATAAATGC AGATTAAGA AGAAGCTATG TCAATATTT
15301 GTTCTTACA TGAGTGGTT AATATATTA AAGCATTTAG CAGAGTGGT
15351 GGCAGTAAAT AGATTTAAA TGTAAAGTAT AGTTATCTCA AATCTCTTG
15401 CTTCAGGAA TTTTTCAGA CAGACCAAG TATGCAAGT TACATATCA
15451 TATATGATA CATGAGATA CATATTATA AAGGAGACT CAGAGAGCA
15501 GCTTATAAAC AATTAAAGC ATAAATGCG ATTATATAA GAGCAGTTT
15551 CCAAGTCTT CTGATCATT GCAACACAG AAAATGTTA TGTTTTCTG
15601 CTTCATTGA GTAAACAGA ATGGAATTC GGAAGCTAT AGAGAGCTT
15651 GTAAAAAAA ATCTTACTT TTAATATAT ATACAATTAT GATGAAAG
15701 CAAAATGCA AGTGTAGGG AAAATATTA ATGTTAATT TATTCAAA
15751 TAAAAGCTT TTTTATTTT TTTTCTTCT TTTTCTTCT TGTAGTCTT
15801 ATCACTCAG CTGAGGCGA GTGCTGTAT CTGAGTCA TACAAGCTC
15851 AGCTGCGAG TTTAGCAAT TCTCTACTT CAGCTTCTG AGTAGCTGG
15901 ATTACAGCA CTGCGAGC AGCTGCTAA TTTTCTTAA TGTTTATTT

15951 TTATTTAGT AATATATCA ATATTTAT TTATGATC TGGATTTTA
16001 GTATACAAA AAGGCAATC TCTATTCAG GCTTTCTCA CCGTACAGC
16051 TAATGCTTC TTAGATTAGA TAAGTCTTG TGTCAAGAT GTGTGATG
16101 TAGGATGTT AGTACATCC CTGACATTA CCGATGAT GTAGTAGAG
16151 TGTATAGT ATAGCAACA TAAATACCT CAGACATTA TGAATGCTC
16201 CAGGCGGCGG AGTGAAGAG CAGTGGCTG TACCGAGTT GTAGAGAAA
16251 TTATTTATG TTTCTGTAG TACTGTATA ATTTCAATAT TTTCAATTT
16301 AATCAGAGA TCTAACTCC ATTTAGATT TATTCCTATA TATGCTGTA
16351 GCTATTCAT TAATTTTCC AATGTTTAT CAGTGTGCT CATCAGCAT
16401 ATTTAAAGT TTATCTTTC AAGTGAATG AGATAAGAT CAGATCTTA
16451 AGCATACAT GTAGTGTAT CTGTTTGA TAAAGTATA TTTGATGTT
16501 CTGCTGAT CTATGATCT ATCTAGCAT GTACAGAGT CAGATCTTT
16551 TAATTAAGA CATTTGCTG CTTTCTTCA CATTAATGA CTTTATTTT
16601 AGAAAGTTT TACCTTGA CAAAATTTA GAGAAAGTA CAGAGCTTC
16651 TCAATTTAC CATGAGCAA AGCTGTATG GTACGCTGT ATCTAAATA
16701 AAGTTGAAA TTTTAAAT ACTAAATA TATTACTCT GTTCAATTT
16751 TTTGTTTTG TTTTCTCTC TCAGTCTCT CAATTATAA TATATTGCA
16801 TTTCTTGGC TGTCTCTAT TTTATTCAT TTTATTTAT AACTTTTCC
16851 TGAGATAAA ATATTAGAT GAGAGAAA AGAATATTT GTCACTTGA
16901 TCTAACTTC AATCATCTT AATTTATTT CCGACATAT CATGAAAGT
16951 ATGTTTTTA TTTGCTGTG TTATCTTTC AGCTTTATC AAAAGTCCCT
17001 TGTAGAGAA AATAAGCAT CTGTGAAAT TAGATCATT TAAAGCTCT
17051 CAAATCAGC AAGATTTGA GCTATTCAT AACATGCT TGTTTTATA
17101 TTTATTTGAG TTTGCTATC CTTTGTAT TGAAGCTAT TTTTCTACC
17151 AGATACATA ATCAGAGAA AAGAAATTC CCGTACAGC GGGTAAAGC
17201 AGATAGAGC TCGAAGAGG AAGTACAGG ATTTCTGGA TATTGCTTT
17251 TCTGCGAGG TAAATCTCT AATTTCTAA GCGTGTCAA GTAGAGTT
17301 AATATGTA GTAGTGGGT AAGTGGAT GGGATGCGA CAGAGATA
17351 AAGCGATTC ACTAAATTA ACTGTCTTT GAATGATGA GAGCTTCAT

17401 GCAATTTGAG ACAAGAGAG AATTCTGCAA CTGTGTGGCT AGAGAGGGCT
17451 TAAAGAGAG TAAAGAGAG AATTCTGCAA CTGTGTGGCT AGAGAGGGCT
17501 ATACATGAT TTAAGGAGAT CATGAAAAA TGGTCAGATG GATAAAGATA
17551 AAAATTATGA TGATAAGGTC CTGGGAAATC TGGGAGTTTG AGAGAGATTT
17601 CTAGGGGCTG TTGATGAGAG GGGCTTTGTC CAAGGCTGTC TTTCCTTATC
17651 TAAAGGAGAT TGTGCTTATG GCTTTGGGCA GAATATGGTT TATAGGAGAT
17701 AATTGTTGAA CTGAATTAAG ATTTAAAGCC CTATTTAAAG CTGTGATTTT
17751 TGGGCTGAAA TCATTATTTG GCTTGTATCT CCAGAGATTT ATAAAGTGGC
17801 AATTATTTTA AAATATTTGT ATTGTACTTT CTAGGATGAA AGTGTAGGCA
17851 GCTTCTGAGA TATGATGTA CACTGTGAG TGAAGCAGAT CTTGTTGGCA
17901 GAGATGAGA CTTTGGAGC AAGCATGTC TGCATGCTTT ACTGGCTGGC
17951 TGTGAAGGCT GAGCATGAG AGAGATGCG GAGGAGGTC AGGGGATGTC
18001 TGGGGAGTGC GTTCTTATC AGTTGGTAA AGTGTGAGCC CTAAATTTTC
18051 CTGTAGTTT TGGGCTGAG ATTTGCTTT ATTTTGGG CTGGTGGCTT
18101 AGTGAGGCTA GTGGCTGAG ATATGCTAG GTGAAGAGA AGAGTAGGTC
18151 TACTTTTCTG TTCTTTCTAA AGAGAGGTC AAATATTTCT CTGTCTTTTC
18201 AGGAAAAA AAAAGTTTA TTATGCTA ATTTGCTGT CATTTGTTTT
18251 CTATGAAAG GTGTGTGAAA TGTCTTATTT CTATTTTCA CTTTGGCTCT
18301 GATGATTTGG AAATGAGGAG TTGATGCTG GGTGTGGCT TAAAGCTTAA
18351 ACAATAGGCT CCAAGTGGAG CTCTCTTCT GAGAGAGGTC AATGATTTAG
18401 TGCATTTT AAAGGCTGAT TTACAGATCT CCGAGGCTCT TGTACCAAT
18451 TTATTTGCTC AGGATGATTT TTACAGTCTA GACATAATAT TGGATGATAT
18501 ATACTATAGT TAAATTTAGC AAATATGAG TGAAGCAGAT TAAATAGTCT
18551 AGACTTTTTT TATGATGTA ATTTATGTC GGGCTGTCT TGGGAGGCT
18601 AATGCTCTAA TACAGGAGAG AGGAGAGAGA CTTGCAATTT GAGTGTAGC
18651 ATATGAGGAG CAATGATAGA GATATGCTG GGTAAAGCA AGAGATAGA
18701 AGAGAGGATC CTAGGCTGAG ATGGAGGTC AGGAGAGCT CTTGATGAT
18751 TGGGCTGACT GAGGATGAG TAAAGGTTAG CAGGAGGTA ACTGTGATCT
18801 CTATCTGCT AGACTTAAAG CATATAGTCT TGTAAATAA GGGAGGTTA

18851 TGTGTTTTC AAGATAAAA TGTGTTCTC ACATAACTCT GGTCAAGGTC
18901 AGAGAGAGAG AGAAATGCTA AGGCAATTC AGAGAGAGAG CAGATAAAAA
18951 AGAGATGAT TCAATATGAA AAGTGAAGTC AATTGAAGA TTTGTAAGC
19001 CAAATTTGAG ATTTAAAAAG TATATGAGAG ATCTGATTTT ATAGGAGAT
19051 AGAGGCTCT TGTACGATA AATGAAGAT TTAATCTATA TAAAGAGAA
19101 TAAATTTGTC AGTAATGAT TTAATTTAT TTTTAACTC AGAAAAATC
19151 TGCATATACA TGTATATAT ATATGATCT GTGTATATAT ATATGATGTA
19201 CAAGATGATA TTTTCTATA TGTATAGCT GTGGAATGAC TAAATGATC
19251 AATGAGAGTC TTGATTAAT CATATGATC ATTTTCTCT GGTAAAGGCA
19301 TTTAAATCT ACCCTGTCAG CAATTTTCAA GTATAGAAAT TTTAGTAAAC
19351 TGCATGAGA TATTGTAGAA TGCATGCTT AAATTTATTC CTGTGCTCTC
19401 ACTGAATTT TGTATGCTT GATTAAGATC CTGTAAATTC CCAATTTCTC
19451 CAGAGGCTCT GGTAAAGGCT GTTCTGCTCT CTGTCTCTT GAGTTTAAAG
19501 TTTAGATTT CAGATGCTA GATGATGTC AATTTGCTCT TGTGTGCTCT
19551 GCTTATTTCA CTTAGGATAA TGTATGCTA ATTCATCTCT GTTGTATAA
19601 ATGAGAGAT ATTTGCTCT TGTATGCTA ATTTGATCT CATTTTATAT
19651 ATATATAGCA TTTTCTCTT ATCCATTTAT CAGTGAATG AGACTTAAAT
19701 TGAATTTCT ATCTGGGCTA TGTGAATAA TGTGCAATG AACATGGA
19751 TGTAGATGTC TCTTGAATG ACTGATTTCA TTTGCTTTG TGTATATGTC
19801 AGAGTGGAA TGTGCTGAT ATATGCTAG TGTATTTTA ATTTTGGAG
19851 GAAATGCTCT ACAAATTTCC ATATGCTCT ACTAATTTAC ATTCAGGCA
19901 AAGTGTATA AGGCTTCTCT TTTTCTGCA TGTGAGGAA CATTTGCTCT
19951 TTTGTAATA ACCATTTCTA TGAAGTGAAG GTGATGCTCT ATATGCTCT
20001 TAAATTTAGT TTTGCTGAT ATATGCTAG TGAAGTGAAG TTTTAAATAC
20051 CTGTGCTCTA TGTATGCTCT CTTTGTAGCA ATGTTATTT AGCTTTTCTCT
20101 CATTTTAAA TGTATTTAT TGTGTTCTCT CTTTGTAAAT GTGTGCTCT
20151 CTATATATTT TGAATTTA ACCCTTATC AGATGATGA TTTGAGAGCA
20201 TGTGCTGCA TGTATTTAGT TGTGCTCTCT CATGTTGAT TGTGCTCTCT
20251 GTTGTGAGA AGCTTTTTC TTTGCTGCA AACCATTTAT CTATTTTCTCT

20301 TGTGTTGAG TATATTTCA GAGTGTATC CAAAAATCA TTTGAGGAA
20351 TAAATGAGAG AAGCTTTCT CTATGTTTT TTTAGTATG TTTATAGTTT
20401 CAGGTCATAT GTTAAATCT TTAATGATTT TTTAGTATG TTTGATATAT
20451 GAGGAGGAT AAGGTCAGC TTTTATCTT CTATGATGTC ATATGAGTTT
20501 TGTGAGGAG CATTTATGA AGATGCTGTC CTTTGAAGC TGTATGTTAC
20551 TGGAGGCTCT GTAGATGAT TGAATATAA TGTGTGGTC TATTTGGA
20601 CTCTTTATC TGTATTTA GTTATATAT CTCTTTTTT AGAGGCTCTA
20651 TGTGTTTTG GTGATGAGC CTCTGATGTC AATTTGAGAT CAGGATGAT
20701 GATGAGTCT AGCTTTGCTC TTTTGTCTA AATTTGCTTT GGTATTTTCA
20751 GTTTTATAT TGTATGAA TTTAGGCTT TTTTTTTT TTTGATGAT
20801 GTGAATATC CATTGAGAT TTTGATGAG ATTTGATGTC ATCTTTGGCT
20851 AGTATGATA TTTTAAAGT ATTAATGCTT CAAATTAAG AACAGAGGCT
20901 ATTTTGAAT TTTGTTTTT TTTAATTTCT TTTGAGGTC TTTTCTCT
20951 AATTTAATG TTTTATTTCC ATAGGTTTT GGTAAAGCT GGTGTTGCT
21001 TATGAGTAA TTTTATGTC GTGATTTGTC AGATTTGAT GAGGATGAT
21051 CTTAAGGAT ATAGATGTA CCAATTTCT AGTCTTCTAT CCGTCACTCT
21101 CTTGAGGCA TTTGAGGCA GTGAGGAG TGTATGAT CATTTGATG
21151 CTTTGTATC CTATAGCTT AGCTGAGCT TATGAGTGA AACATATAAT
21201 GTTGTCTCT CCAATTTCTA GTTATTTCT TGAATATTT GGTGCTCAAT
21251 TGTATGAGA TGTGAGGAA TGTGTTAT TTTTCTCT TGTATGCTA
21301 GTAGTATTT ATAGTATATA CATGAGGAA TTTTCTAT CATTTGAT
21351 TGTAGGAGT TGTAGGCT TGTATGCT TGAATTTGTC AATTTGCTCT
21401 CTAGAGGAT GAGGTCAGC GTGTTTTTT CATATATGA CTTGCTCTCT
21451 TGTGAGTGA TGGGCTGAG TGGGATGCT GATCAAGTC GTAGTTCTAC
21501 TTTTATTTCT TTAAGGATC TGTAGGCT TTTGATGAT GGTGATCTA
21551 GTTATGATC CAGGAGAG TGTAGAGTC TTTGCTGTC ACTGATGCA
21601 CAGGATGAT TATTTATTT TGAATTTTT ATATGAGCA TTTGAGGTC
21651 AGTAAAGTGT TATGAGTGT TGTGTTGAT TGTATTTCT CTGATGATTA
21701 GTGATGTTGA GCAATTTTT ATATATTTGT TGGGATTTG TATATTTCT

21751 TTTGAGGAT GTTATTTCT GTGTTTTCT CATTTTTGA TGGGATTTT
21801 TTTTTTTTCT TGTATATTT GAGTCTCTC TAAATTTGTC ATATGAGGCT
21851 TTTGTTGAT GTTATGTTG TGAAGATTT CTTGAGTCT TTTGTTGTC
21901 TTTTACTCT GTGATTTAT TTTTCTGTC TGAAGGATC TTTTATTTA
21951 ATTAAGTCTC AGTATTTAT CTTTCTGTC TGTGTTTTT TTTGTTGCT
22001 TTTGTTTTG CTTGTTTTG CATGCTCTT TGTGTTTTG GTGATGAT
22051 CTTTGTCTA GGTATATCT AGAGGTTTT TGTGTTTTT CTAGATTTT
22101 TATGTTGAG GTTATGAT TAAAGTCTC ATGATCTCT AGTTGATTT
22151 TGTATAGGT GAGAGATGAG GATGAGTTT CATGCTCTA CATGTTGCT
22201 GGTATATC CAGTATGAT TTTGGAATA GGTATATAT TTAAGGCTT
22251 ATATATTTAG GTTGTCTAT TTTGAGTGA TATTTATTA CAATATCAT
22301 ATGCTCTGA TGTATGAGC CTTTCTCTA TATATATG TTTTCTCTC
22351 TTTTCTTACA GTTTTTTCT TAAAGGCTA TTTGCTGAT AAGGTTGAG
22401 CTAGTTTTG TGTTTTTG TTTCTATTT CATGATAT TTTTCTGAA
22451 CTTTGTGAT TGTATGAT TGTGTTTTA AGATGAAAT GAGATGCTCT
22501 AGGAGATAT CTTGTTGCT TTTTATTTT ATTCATTTAG CAGGCTTTT
22551 GATTAAGAA TTTAATTTCT TTTATTTCA GGTATATAT GAGGAGAGC
22601 GATTTACTC TGTATTTCT TTAATTTTT TTTGATTTT TATAGATCT
22651 TTTTGTCTT TGTATGCTC TTTTCTCTT CTTTGTGAT TGTGCTCTT
22701 TGTATGTC TGTATTTCA TTTTATTTT TATCTTTT TGTGCTCTCT
22751 ATAGTTTTT CTTTGTCTT TACATGAG GTTATATA GATATGTTT
22801 AAGGCTAT TTTAAGTGA TACAGGCTA ACTTTGAGC CTTAAAAA
22851 CTATAGCTT TTTATGCTA AAGGCTCT CATTATAT CTTGATGTC
22901 ATATTTTAC TGTTTTTGA GATGTTTTT CTTATGCT ATGCTTAAAC
22951 AATATTTCT AGGAGGTC ATTTTAAAT GTTTGCTCT TTAATTTAT
23001 AGTATGATA GATTAATTA ACATAGGCT ACTATTTAT TGGGATTTT
23051 TAAATGATC ATGATTTAT CTTTATGAT GAGATTTTT TTTTCAATTT
23101 TGTGTTTTT AATGATTT CTTTATTTT AAGTGTGAG ATTCAGATTA
23151 GCAATTTTTT TAAAGGCT CTAGTATGTC TGAAGGCT CAACTTTTTT

23201 TTATCTGGAG ATCTCTTAC CTCTCTTCA TTTTGAATA TAACTTTTGT
23251 TCCATGATTC AATGAGCAA AATCTTTTTT TTAATTAATC AAGTCCAG
23301 GGTAAAGACA ATTAATCTTT TTTTTTTTT CTGAGACCA GTTCACTCT
23351 TCTTGGCAG GCTGGAGTC AGTGGGCAA TCTCTAGCT TACCGAAGC
23401 TCTGCTGTC AGCTTCAAGC GATCTCTCTG CCTCAGCCTT CTTGAGTAGC
23451 TGGGATTACA GCAATGCAAC AGCATGCTCC GCTAATTTTG CATTTTATGT
23501 AGAGAGGGGG TTCTGCAATG TTGCTCAGCC TGGCTTTGAA CACCGACCT
23551 CAGATGATTC GCGCAGCTAG GCGTCCGAAA CTCTCTGGAT TGCAGCTGTG
23601 AGCAGCTGCG CCGTCCGAGA ATTAATCTTT TTAATCTTCA GTTTCAGCAA
23651 CAAGAGATTC AATTAATAAA TTTCACATTA CTAATGCGC AAGGCTGCA
23701 TTCAAAATAA GTATTCAGAA AATCATATAA AAGCAGATA AGATAAGAG
23751 ACTAAATATA TGCAGTCAAT TTATGGAAGC AATCTGACTA GATTGATTC
23801 AGACTAGETA GGAATCAAAAT TTAATAAAAA CTTAATCTTT CTTCAGCTTA
23851 TAAACTTTAA AGCTGCTTTG TGGAGCAAGT TCTTTTATC TGTGGGAAA
23901 GATCTGAGT AAGTCTGATA GAGTCTCTAT TCAATTAAT CACAGAGCA
23951 ATCTTAGCT AGTAATTAAT CTATCTGCGC CAGTGTAAAT CTGAACCTT
24001 CAATACTTA TGCAGTGAAG CTCTCTTTG CATCCAGCT TGGTACTTCT
24051 TTGCTCTAG AAGCAGCAG TGGTTATCA TCGACTTAT CTACTGACT
24101 AGCTGCCAA TACCGAGTAG CTCTCTTTG TGGCCTGTC AGGAATGCT
24151 TTAGAGGAA AGGGGATAG GAGTAAGGG CTGCTACTAT TGTGATCAG
24201 CCAAGGGCT TGTGATAT TCGATGCTTC CTTTCTCTC AAGAGCAAG
24251 TCGCTTCTT GAGAGCTCTC TCACTAGAAC TTTCAGAGG TCAATGAGG
24301 CACAGAGAA TAACTTCTCT TACCGAGCT CTTTTCAGG CTGCTCCAG
24351 AGCTTTGCT CTGCTGCTT AATGCTTTA AGGACAGCT TGCATCTCT
24401 TCGCTTCTT CTGCTGCTT CCGTCTGCT TCTCTCTCT CTGAGTATA
24451 GCGTTTCAA TCACTCTCT ACTGCCAAA CTCCAGAGG CACAAGTCA
24501 ATCATCTAG TGAATCTCT GAGGCTCTT GTTAAAGAT GGGAGAGCAG
24551 CCGTCTCTT CCAATGAGCT CTGCAATTC AAGACAGCT TAAATATTT
24601 TTCTCTCAA AATCTTAGG CCGTCTCTT TTAATCTTC GCAATTTTA

24651 TGTATTATTA CTTTATATCA TGAGTANGA GTTACAAAC TGGTTTTAG
24701 AATCTGCTT AGGAATGTT TTACTGCTAG TTAGCAGCT CACTTTATA
24751 TANGATATA TGATATATTT CTTTGGTTC TGTGCTCTG GAGCTGAGC
24801 TCACTCTAG GAGAGAGTC CCAATTTAAG ATTCTGTTAC ATAAAGCAGT
24851 GCAAAATGC CTTTAACTTC AGGTAATAA TTACAGGAA CAAGAGAAA
24901 ACAGAAAAA AGTAACTGG TTATGATATC TGAGTGGCTT CCGTCCCTCA
24951 TCTCAGAGC CAGCAGCTGG GTGAGATCTC GTACAGGACA ATCTGATCA
25001 AGGAGAGCT CCGATTTGAT CCGCAGTCC CCGTCAATTC CAGAGATCTC
25051 AGCAAGCCAG TTACTTTCCT AGATGATTC ACATGCTCT CAGCTCTTTA
25101 CATCTTTTT CTAGCAGCT CTTAGAGCT ATGCAATCT GAGAGCACA
25151 GTACAAAGA TTAGTAGTC TCTTAGACT TGGAGAGTC AAGAGATAAT
25201 GCTACATCT GACTTAGCT TTATCAGCTA TGGAGAGTC AGAGATAAT
25251 GCTTGTCTA GAGATGAAT TCAATGACT TCCAAAGGC ACATAGCAG
25301 TTGAGGAAA GCTAAGGCA GATGCAAT CTCTGAAAT CAGGCGAGG
25351 GTCTCTTCA TTGAGGACA TCAATCTAA GATAATCTTT GTTGGCTGA
25401 GTTTCAGAG CAGCTGAAG TTCAATGAAA ATAGAGCAG CATCTTTATC
25451 TGAAGAGCA AGGGGATCT TGGCTCTAT CATCAATA TCACTTAT
25501 AATATACAA CATTAATAG TTAATATAGA GCGTTCAGC CCAATATCTC
25551 ATTTTCTCC TTGAAATCA ATGTAAGAG ATCTTATAC ATGATTTAC
25601 AGTCTCTCA ACATTTTAA GTACTTCAA TGTGCGCAA AATCAGAG
25651 CAGCGGCAAT CTGAGATCA CATTAATCA TGTGAGGACA GTAGAGCTT
25701 GTGAGAGAG CCGTCTCTG CAGGCTAGAG TTCTTGGAA CAACAGCT
25751 TTCTGAGAG CCGTATAGG AAGCAGAGG GTCATCTAG ACATATATC
25801 TGAATCAAT TCTATTAAT TCAATCTTA GGAAGCAAG CAACAGATT
25851 GTTCTGACA AAGAGTACA GCGTCTACT GTAACTTTG TGAAGAGCC
25901 AGAATTAAT TCTGAGCT AGAATATTT CTGAAAGCA AATAAGCTC
25951 ACATCTCTC TCGTTCTTT TCACTCTCT TTCTGCCAA ACACATGGA
26001 TATTTGCCA AATCTGAGC TTTCATATG TGAATGAGC CAATGAAAT
26051 TTCTCATGG ATCTGATCA CAGATGACA GTTCTCTCT TGTCTCTCT

26101 CCGTTTCTC TCAAGACAGA GTCTGCTAT GTAGGCCAG CTGAGTACA
26151 GTGGCTAAT CTGGCTCAC TGCAGCTCT GCGTCCAGG TTAAGCAGT
26201 TCTCTGCTC CAGCTCCCG AGTAGCTGG ATTAGAGTG CACAGCAGC
26251 CTGGCAATTT TTGTATTTT TATTAAGAT GGGTTTCA CATTCTGCG
26301 AGGCTAGCT CAGCTCTCT ATCTGAGAG CAGGCTCTCT CAGGCTCCCA
26351 AAGGCTGGG ACTAGAGCA TGGAGCTTC CAGGCTCCA GTTCTGCT
26401 TTTATAGCTA AATGCTTCC AGGAGCTCT AATAGTCA TAAATGAT
26451 TTAGGCCAG CAGAGTGGT GAGGATATA ATCCAAAT TTTGAGCAG
26501 CAGGCTGGA AGACTGCTT AAGTAGAGC TGTGAGCTA GCGTGGCAA
26551 CATAGGAGA CCGTCTCTT ACAAAGAAA AAGAGAGCA GATAGCCAG
26601 CATGCTTTC CATGCTGTA TTCTGCTTA CTGAGGAGC TGAAGCAGCA
26651 GCACTAGTC AGCTGAGAG TTCAAGCTTA CCGTGGCAA TCTTCAAGC
26701 ACTGCTCTC AGCTGATTC AGAGGCCAG CCGTCTCTT AAGCAAAAC
26751 AAAAAAGAAA TATTAAGTA ATTTCCAA CAGAGAGAA ATATAGCAT
26801 GCTTATCAC TTGATATCA CAGCAAGAG TACTTAAGAT AGAGTATCA
26851 ATTCAGTAA TTGCTCTCT GAAAGCTAG GTGCAAGCC AAGGCTCAT
26901 TTCTAGGTC CCGTCTCTG GTGCTATCA CTACAGAGG CAGAGATTC
26951 CAGGAGCTA CCGTCTCTT TCAAGAGCA AAGCTCTCT TAAAGAGCA
27001 GTAGGCCAG ACTGCTCTT TCTGCTGAG TCTCTTTAT TTCTCTCTT
27051 TCTTAGGAT CAGGCTCTT CTAGTATTT GCGCTCTCA CACAGAGCT
27101 GCTGCTGGA AAGAGGAAA GGTATGATC TCTCTCTAC ATAAATCTT
27151 CCAAGAGTA ATCTGCTCA AGTCACTTT TGTAGCTAA GCAAGAGAT
27201 GCTATATAA TTAAGGAAA TCAACAAAT TAAACAAAA TAAACATAA
27251 AGGCAAGAA AATGAAAGC TATCTATCT TCTGAAACA CTCTGAGCT
27301 GTATAGCTA TTCTTTCTT GTAGGAGCT AAGGTTAGG ATCTATCTT
27351 TGTACAGGA AGCTGAGTA TATGCTCTG TAAATATGG CAGCATCTC
27401 ATTTGACTT GATTTCTAG TGGATGACA TCAATTTCT AGTGAATGA
27451 TCTCATAGT GAAATATA CCACTTCCA TACTATTTT GTTGGCTGG
27501 GTATCAGAA AATGAGCTT GAGGCTCAA GCTCTGAGG TGTGAGGTA

27551 CAGTGGCAG TGTAGAGCC AGGGGCCAG TADGCTACT GAGCAGTCA
27601 CATGAGCCA GTTGAATTC AGTGTCTCT TAACTTTAA ATAGCTGCT
27651 GATTTGAGG ACATAGTACC CTAAGAAAT GTGAAGCAT TCGTTTATG
27701 AATTAATTA ATTAATTACA GGTGGAATG GTAAATTTT GTTAAATAA
27751 CTCTATTAG ATTAAGCTCA CTTTAAAA ATGTGAGCAG CAGAGCATTT
27801 TAAATGAGC ATGATGATCA CATTAATTT CTATGATCG GTGCTAGCT
27851 GTAGGAGAG AATGCTTTC ATGTTGTTT GGGGAGCTG TTGGGCTCT
27901 CCGTCAATTT CAGGCTCTT ACCGCTTGG GTTCTCTAG GAGAACTCT
27951 ATCAGAGCA CCGTATGCG TACTTACAT TCTGAGCTG ATCAGGCTCA
28001 GAGCAATTT AGTCTCTCA AAGTACCAA AGATTTTAC TTGAGAGTAC
28051 TTATTTCTT TCACTCTCT AGCTCTATC ATTCTTCCG GCAAGCTAG
28101 ATCTGCTGC CATTTGAGC CCAAGAGAT CAGTATGTT TACAAGGAG
28151 AATGATTC TGTGAGCAT CTTTCTTGG CATGCTCAA AAGAGCTCC
28201 ACATGTAAG CTACTCTCA TACTCTAAT GCACTGAGC TTCAAGGCA
28251 GAGAAATAT TTATAGTA ACTCAAGT CCGCTCTCT ATAGGAGAG
28301 CATCATCTG CAGGCTCCA GCTCAATTC TCACTGAG CAAATTAAGT
28351 CTAAAGTCA AAGTTTCAA TGGATTTGG TCAAAAAAT ATCACTTAC
28401 TGTGACTTC AGCTCTCTG TACTAGTAT TTACTATAG CAGAGAGAG
28451 ATCAATTTT CAGTATCAC TTCTTTCCG TCTCTCTCT AGCAAGTCA
28501 TTGGGAGGA GTTGGGATG ATGAGTTAA AGGTAAGCAT TGGTTGAT
28551 CCGTCTCACT TCAAGTCA CAGAGAGCC AGGAGGCTT TACTTTCCG
28601 CAAGCATTT ATCTCAAGC CCAAGAGAG GATGATTTT CAGTGAAGA
28651 AACTCTCTA ATCTGAGT TCAAGGACA ATGATTAAG CTACTTTCT
28701 TTGGAAGTT AATTTAGAG CTAATGATC AAGCAGATG AAGGAGTCA
28751 ATGATGCT GAGGATTTG AGGTTGCTG GATAGGCTC TGTGAGGA
28801 GATCAAAAT CATTTCTAG TACAGAGCT GTCAGTACA TCTTTTCTA
28851 TATAACTTG GAGGATTTT AGATCTTTT TGTAAAGCT TCACTACTAT
28901 TAACTGCTA TACAGGATA GACTTTCTA TATTTCTCT TTTTTTAA
28951 ATAGTTTCA GAATATGCA AGTAAGAT GATGATAGC TCACTGCAA

29001 AAATGCGA CACTAGAAA TCATGTAGA TAAAAATTT AAATCTACT
29051 TCAATTAGC GACATTCAT GCGCTGACC ATCTACTGC TTTCCTAAA
29101 AACAGATAA TTGCTGTGC ATTCTTCAG ACTTTTCTC ATACATTTA
29151 TATGTAGAA TGTAGCAATG TATTTGTATA GATGTGATCA TTCTATATT
29201 GTTATTGAT TTCTGACTT AATAAAATY GACCTTATC CTTATCATC
29251 GTTATGTGA TTCTGTATA TCAATGACT ATAAATTAAT TAACATATTT
29301 CTTATTTGG CATTTAATY ATTCTACTY TAAAAACAT GTTCTCAAT
29351 GCGAAGAAA GCGAAATTC ACAAATGGA TCTAATTAAT CTAAGAGCT
29401 TCTGACAGC AAAACAATY ACCATGACG TCAATGGGA GCGTACAGAA
29451 TCGAGAGAA TTTTTCAGC CTACTCATY GACAAAGGC TAATATCCAG
29501 AATGTAGAT CAATGAAAC AAATGTAGA GAAAAAGCA ACCCATCAA
29551 AAGTGGGTG AAGGATATG ACAGACACTY CTCAAAAGA GACATTTAG
29601 CAGCAAAAG ACAGATCAA AAATGCTAT GCTCACTGC CATCAGAGAA
29651 ATGCAATCA AAGCCAAAT GAGATAGCAT CTGACAGAG TTAAATGGC
29701 AATGATTAAT AAGTACAGAA ACAGAGGTG CTGACAGAG TGTGAGAAA
29751 TAGGAGACT TTTACACTY TGTGTGAGC AGAATCACTY GACCCGGGA
29801 GGGGAGGTT GCACTGAGC GAGTGGGCG CAATGCACTY GAGCTGGGC
29851 GACAGAGCA GTATGCAAT TCAAAAAA AAAAAAGCA CACCAACTY
29901 CTCATCTTA ATGTTGTAT CTATGTGTA TCTTGCATA TGTCTGAG
29951 ACAGAGTAT CTTTCTGCA TATGATCTA CAGTATTTT TCTTATAGC
30001 ATTATATCT CATTAAATG AGCAACAGAA ATGCAAAAG ACAGTCAAT
30051 TCTCCCTTC CATGACCTAA TTTGCTTCA CTCTTGCAT ATCACTATA
30101 ACATGATCAT TCTGAAATY ATCTACCTA AATCTATATA TAAAAAATC
30151 CTTCCCTTGA ATTCAGATC CTGAGAGCA AAGACGAGC TCTAAAGCA
30201 AATTTGTTA ACACTGAGC AGTGTGCTG TGTGACTTTC CATTTTGTA
30251 CTAATTTGTC AGCTGTATA CCAATATCA CCAAGTAAA CAATATTTCC
30301 TTGTTTTTT CTGCTACAA CCAAAATAA TTACAAAGAT CAATAAAGT
30351 AAAATCTAA AATAACTGAC TTTCTGTATA TATCTGCTC TTGCTGAAA
30401 AATGGGTAG GTTACTTCTY TAAAGCAGT CATGATTAAT TGTACTGAT

30451 ACAATATTA GGTGTGACA TACTAGTAT AATTTCTGT GTCTGTGGG
30501 TCTTACCTAT TTGGGTCAA AATAAGAGG TTTATTAAG TTATTAATAT
30551 TCAATTTAT TATCTTCTT AACAATATG TTGCTGTA GTTCTATTC
30601 CAATAATTA TTTGTAGCT TGGAGGTGC TTCTAAATY CTGTATATT
30651 TTTATATTC AATTTACTY TAAATATTT TAAAAAGG GTCTGTAAA
30701 TTTCTAATA ATTATATAT TATGTTTTY TCACTGATY TTTGTAAAT
30751 GAAAGCCCT AAAAAATGA AATCAATTT TGAATATG TGGACAGAC
30801 AATTTGTTA AATAAGAGA CAGAAAGAG GCAATATCA GAGATAATA
30851 TTCAATATC CTTATATTC TGTACAGAT TTTATAGCA AGTGTGCAA
30901 AATTTGATA TCAATATAAT GATAACAGT TCAGAAAGC ATTCTTTAT
30951 CCGTAACCT TCAATATGA AACTTTGTA GTAGGAAGT AGGCAAGCA
31001 TATATTCCT TTGAAGGTG CAGAAAAATG TCATGCAAT TCACATGCT
31051 ACTCTTCAAG CTTAAAAAA ATGCACTGCA AAGATTTAG AACATAGCA
31101 TATTTATGG GTACCTTAT GTTACATAA ATATTAAGA TATCTCAT
31151 ACCTCTTCA ATCAGATTA CTGCTGACA TTTATGACC ACTTTCTATG
31201 GGGAAAA

FEATURES:

Start: 2191
Exon: 2191-2367
Intron: 2368-8318
Exon: 8319-8460
Intron: 8461-9761
Exon: 9762-9806
Intron: 9807-11566
Exon: 11567-11594
Intron: 11595-14298
Exon: 14299-14426
Intron: 14427-14509

Exon: 14510-14564
Intron: 14565-17152
Exon: 17153-17259
Intron: 17260-17834
Exon: 17835-18025
Intron: 18026-24959
Exon: 24960-25093
Intron: 25094-27056
Exon: 27057-27121
Intron: 27122-27913
Exon: 27914-27996
Intron: 27997-28492
Exon: 28493-28564
Stop: 28565

CHROMOSOME MAP POSITION:

Chromosome 1

ALLELIC VARIANTS (SNPs):

DNA

| Position | Major | Minor |
|----------|-------|-------|
| 267 | T | C |
| 284 | G | A |
| 1269 | T | C |
| 2487 | T | C G |
| 4486 | G | A |
| 4522 | G | A |
| 4523 | C | A |
| 5075 | T | G C |
| 5450 | T | C |

| | | |
|-------|---|-----|
| 5450 | T | C |
| 5995 | G | A |
| 6241 | G | A |
| 8479 | C | T |
| 10045 | C | A |
| 10045 | G | A |
| 11994 | G | A |
| 14070 | A | G T |
| 15635 | T | C |
| 17618 | C | T A |
| 18520 | A | - C |
| 18526 | - | T A |
| 18525 | - | G A |
| 19189 | T | C A |
| 19259 | C | T |
| 19325 | G | T |
| 19346 | G | T |
| 20845 | - | T |
| 20845 | T | C |
| 22234 | T | C |
| 22234 | G | T |
| 22247 | C | T |
| 22334 | A | G |
| 23033 | T | |
| 23036 | - | A |
| 23421 | A | G |
| 25582 | T | C |
| 26407 | C | A |
| 26473 | C | T |

26844 G A
28384 A C
28417 A C
29265 A G
29484 A G
30417 T
30783 C G

Context:

DKA

Prallie9

267 CCAGCCTCTCTTAGGCTCTAAATATAGTCCAAAAGTTCCAGAGTTCTTTGTTACCCA
TGAAGCACATGGAAGGTGCTGACAGGGCAACTGGCCCTGGAGCAGAGAGTAACCTG
CATAGAACTGTCCAGCCTCAGAGGAGTCAACCCAGCAGAGAACCTGGTGGGAGTA
GGTGAAGCCAGGGTTCCAGGCTCTGACCCCTGCCAAGAGAACTCAATTAGAAGGTCAACA
AOCACACATACTATTCTGGGTCTCA
(T, C)
GAGAGCCAGGGAAGGGAAGGCAAGATATCACAAGCTGAAGTTTGAAGCTCTGGGC
AGAGCATGCAATCTAGGCTTTTGGCCCTACACCATGCAATCATATGAGGGCATCATAC
AACCATCATGATTGGGGAGGAATAGGCAATAGAGCAATCATATGAAAAGCTCAATGC
CATGAGTTACCCAGAGAGGCTGTGTAGGCAAGGATTTCTGAGACCCCTGTCAATAACA
ACATCTAGTTGAAGTTGGAGTTAGCTAGGAGTAGGCAAGTCTGGGAAGAGAGGCTG

284 CCAGCCTCTCTTAGGCTCTAAATATAGTCCAAAAGTTCCAGAGTTCTTTGTTACCCA
TGAAGCACATGGAAGGTGCTGACAGGGCAACTGGCCCTGGAGCAGAGAGTAACCTG
CATAGAACTGTCCAGCCTCAGAGGAGTCAACCCAGCAGAGAACCTGGTGGGAGTA
GGTGAAGCCAGGGTTCCAGGCTCTGACCCCTGCCAAGAGAACTCAATTAGAAGGTCAACA
AOCACACATACTATTCTGGGTCTCATGAAGAACCCAGGAGC

ACTTATAGTTGAGAAATATGCAAAATCTATCCCAATGGTTGGACCCCTAGTCCATT

4488 TGTATGTATCTCTACTGTCTCATGAATACTATGTGCTGTGTGTTTAAITGAATGTT
TTGGCATGCTGTCAAAAATCAATTTGACCATAAATGTCAAGCTCTATTCTGAGTCTTCA
ATTCTAATCCATTGATCTATATGTCTATCTAATCTCATGACACAGAGAGTGAAGGATG
GTTACCAAGGCTGGGAGGATAGAGGGAGCTGGGGAGGAGGTAGGGAAGGTTAATGG
GTACAAAAAAATAGAAAGATGA
(G, A)
TAACACCTACTATTTCATAGCATAGCAGGCTGGCTATAGTCAATATAGTGTACACTTT
TAAATAAGAGTGTAAATAGGATTTGTTGCACTCAATGCAATAATGCTTGAAGGCAATGG
TACCCCAATCTTCATGATGTGCTATTTCAGATTGCAATGGCTGTATCAAAAACATCTCAT
TTACTCCATAAATATATACAGCTACTATGTATCCAGAGTATTAATAATTAATAAAT
AAATTAATAGCTATGCTATGCTAGTACCAGACTGCTTACTGTTGCTTTGTAGTAGC

4522 TGTATGTATCTCTACTGTCTCATGAATACTATGTGCTGTGTGTTTAAITGAATGTT
TTGGCATGCTGTCAAAAATCAATTTGACCATAAATGTCAAGTCTATTCTGAGTCTTCA
ATTCTAATCCATTGATCTATATGTCTATCTAATCTCATGACACAGAGTGAAGGATG
GTTACCAAGGCTGGGAGGATAGAGGGAGCTGGGGAGGAGGTAGGGAAGGTTAATGG
GTACAAAAAAATAGAAAGATGAATAACACCTACTATTTCATAGCATAGCAGGCTGGCT
(G, A)
TAGTCAATAAATAGTGTACACTTTTAAATAGAGTGTAAATAGGATTTGTTGCACTCAA
TGGATAAATGCTTGAAGGGAATGGGTACCCATTCTTCATGATGTGCTATTTCACATTGC
ATGCTGTATCAAAAACATCTCAATTTACTCCATAAATATATACACCTACTATGTATCCAC
AAGTATTAAATTAATAAATAAATTAATAGCTATGCTTATGCTAGTACACACTG
CCTTACTGTTGCTTTGTAGTAAGCTTTGAAATGAGGAGTATGAGTCCCGGCACTTTGG

4522 TGTATGTATCTCTACTGTCTCATGAATACTATGTGCTGTGTGTTTAAITGAATGTT
TTGGCATGCTGTCAAAAATCAATTTGACCATAAATGTCAAGTCTATTCTGAGTCTTCA
ATTCTAATCCATTGATCTATATGTCTATCTAATCTCATGACACAGAGTGAAGGATG

(G, A)

GACAGGCAAGATATCACAAGCTGAAGTTTCAAGTCTGGGCAAGCATGGAATCTGAG
TCTTTGGCCCTACACCATGGAATCATATGAGGGCAATCATACACCATCATGATTTGGG
GGAGGAATAGGCAATAGAGGAATCATATGAAAAGCTCAATGCCATGAGTTACCCAGAG
AAGCTGTATAGGCAAGGATTTCTGAGACCCCTGTCAATAACACCATCTAGTTCAAGCTT
GGAGTTAGCTAGAGTAGGCAAGTCTGGGAAGAGAGCTGAACACTTGTCTGTGT

1259 CTTGTCTTAATCACTTACCCCAATAAATCTGGCTGAGAGAGTGAAGGTAGGCTT
AAGCAATTTGGGGCCGAAGCCGGGCAAGCTGGGGAGGCAACTCATAGGCAACAGG
CAATGTAGCAGAAATTTGGTTTATTGTTGAGAGCTGTCAATGAACATTAACAATATGC
TGTCTTAGGCTTAAATCAATGAATAATGAATGAATAAATAAATGAATGAATGTGGGAA
TGGCTATAAGATTTGCTGGGACAGGAGGTGGGGGAGACACAGCTTGGGAAGTCAGGC
(T, C)
TGTAGATCCTAGTTCAACCATGATAGCTTACAAATCTAAACCATCACTTTEAAAT
ATTTTTACTACATTTTCTGTTATCTGTACTGGAGTTTATTAATGTTTCTGGCATGAG
GTAGGCTTCTAGGCAATGAGACCCAGAGGCAACAGAGCTCTGAACCCAGAGAGC
ATATGCTGGTTTAAATGCTGTCTATCTTGAATTTGTTAATAAGTTTTTATCCGGAAT
TTCAATTTGCACTGAGATTAATAATTAATAGCAGGCTTCACTGADCTGTATAGTGG

2487 AGCCATGGAAATTCCTGCTGAGAGGGGCTGGGGGGGGCTTTTACCTGGGCTGT
GTTCTGCTGGGCTGGGGCTGCTGAGGCAATTAAGCTGTACCTGGGAGGAGGGCT
GCTGGGGAGCTGGGGGCTTTCCAGGGGGGGGACCCACTGTTCTTGGGACCCAGAA
GCTAAATGGAGGAAAGAGGTTACAAAGAGGAGGAGAGGGGGGGGAGGAGATGGGG
AGAGAGGGGAGGGGAGAGAGAGGAGGCTTTCTTCAATGCTGGGAGCCCTGGGCTT
(T, C, G)
CAGGGGCTTTCCAGGGGGGCTGGGCTTACCATCAATTTTCTGTTCTGAGAAAT
TGTCTTCCGAGGGGCAAGGGAAGGTCAGAAAGAGAGAGCTTTGGGGCTGGGAG
GAGCTATTAAAGAGCTGAATATGGAAGAGAGGAGGCTGAATCAAGTCTGTCTC
TCAATGCTTCAAGAGGCTTCCAGATGTGTGCTTAAAAATAGCATGTTATCTAAATA

GTTACCAAGGCTGGGAGGATAGAGGGAGCTGGGGAGGAGTAGGGAAGGTTAATGG
GTACAAAAAAATAGAAAGATGAATAACACCTACTATTTCATAGCATAGCAGGCTGGCT
(C, A)
TAGTCAATAATAGTGTACACTTTTAAATAAGAGTGTAAATAGGATTTGTTGCACTCAA
TGGATAATGCTTGAAGGGAATGGTACCCCAATCTTCATGATGTGCTATTTCACATTGC
ATGCTGTATCAAAAACATCTCAATTTACTCCATAAATATATACACCTACTATGTATCCAC
AAGTATTAAAAATTAATAAATAAATTAATAGCTATGCTTATGCTAGTACCAACTG
CCTTACTGTTGCTTTGTAGTAGCTTTGAAATCAGGAAGTATGAGTCCCGGCACTTTGG

5075 TTTGTAGTAACTTTGAAATAGGAAGTATGAGTCCCGGCACTTTGTTATTTCCAGA
TTATTTTGGCTTTTGAATGCTTGAATTTCTATACAAATTTAGACTCAGGCTATCAAT
TCTACAGGAAGCAGCTAGGTTTGTGCTTGGCAATGCACTGAATCTGTAGATCAGTTTG
GGGATTAATCCATCTTAAGAAATATTAGTCTTCTGATCATGAGCAGAGAGGCTTTC
GCTTAGTTAGTCACTTTAAATTTTGTGCTTTTGTGTTTTCCTTTTTCAGCAGAGT
(T, G, C)
CTGCTGTGTGGGAGGCTGAGTGCAGTCAAGCAATCTGGCTCACTGCAAGCTCCGGC
TGTGGATTCAGGGAATTCCTGCTGCTCAGGCTCCCAAGCAGCTGGCACTACAGGCAT
GCCACCAAGCAACTAAATTTTGTATTTTCACTAGAGAGGGGTTTACCATATTGGCCA
GGCTAGTCTGCACTCTGAGCTGTATCCAGGCTCAGGCTCCCAAGTCTGGGA
TTACAGGCTGAGGCAAGCACTCCGGCTTTCTTAAATTTTAAAGGATGTTTGTAT

5450 GATTCCTGCTCAGGCTCCCAAGCAGCTGGGACTACAGGCATGCAACCAACCAAC
TAAATTTTGTATTTTCACTAGAGAGGGGTTTACCATATTGGGAGGCTAGTCTGCAAC
TCTGAGCTGTGATGACCGGGCTCAGGCTCCCAAGGCTGGGATTAAGGGGTGAGC
CAGCACTCCGGGCTTTCTTAAATTTTAAAGATGTTTGTATTTTCAAGGATAC
ATCTTCAATTTCTTTGTAATTTATTTGTTTGTCTTTTAAATTTCAATTCAGACTA
(T, C)
TTATTCATTCATAGTGTTTTGAAGTCCCAATTCCTCTTCACTGTCACTAAGTTTTTT
TTTTCTGTTTTTGAAGGTTTCTATCAGAAATTTGAGATCAGAGTGAAGGATATCTCA

TTCACTCTTACTCTTAGTAATACAGGTCTTCACAACTGCAGAAATATCTCTCCCTTGG
AGCTTTGAAAGATGCAGCTCTGTGGTTATATATCTCTCTCCCTGCAGAAATATCTCTCAAAAGAG
GCTTGGCTTACCAATTCAGCACTAAAATAGCACTCTAGTAACTCTCTATCTCCAAAGCTAT
TATATATATCTGGGCTTATCACTCTCTCAGCACTATAGTGTATACTCTTCTTGGTTCTTC
(5, A)
TTTATATGCAAGCACTAACTACAATATAAAATCTGTGCAGAGTAGCACTCTTCTTGTTCGCA
CATAAAGCTAGTCAGTCACTAGTCTCTGTGCATTAATAGTGTCTCAATAAATGCTTTGT
TTGAATGCATAAATATATAGTGTCTCAGAAAATTTATTTATTCAAAGATCAATTTACTG
CATAGAATAGCGCAGCTGTCTTGACATTTATTCGAATAGCGCAAGATATGGCACTAGCATG
TACATATGCAGTGTGTGTGTGTATGTGTGTGTGCATGTGCATGTGTACTGTGATGTACT

8479 AAGCATGTGTATGTCACTTGCAGAAAGTCTCAGGCTCTCTGCTTGTGTGCACCTATCA
GCTCTGCAACTCACTCTCTCTTATAGAGAGGAGAGCTGCAGCTTGGCTGGCTAATTAC
TTTAACTTTTTTTCAGCTAGTATTAATCAGGATGATTAACATGAGAGTGTGAGAAATTA
TTGAAAAATAGCTCTGGCTTGGCTTCTGTGATTTGGGCTTTTCAGGCAATTTTCTGTGA
TGTATGAGCGAGCACTATGCAGAGCACTTCTGAGCAAGCAGGTAGAGAGAGGGGGAAG
(C, T)
TTTGGGAGCTATCTCTCTAGAGTGAATGCAATAAAGGCAATGGCAAGATTCGAAAG
AAGATTTGTTTTGGGGCTTTAGAGAGCAGCAGCAAGATAGGGAGGTGCAGGTTTC
CTAGCAATAGTGAAGGGAATTCGATATCTCTCCGCACTCTCTCTTCTTGTAGGTATGC
ATGGGAGCTGTGAAGTGTATTAATTAAGGCTAGTGTGCAATAGCAAGCTTGCAGAGC
AGGCTTGGCTTGGCTCTGTGGCTTTTATCACTTCACTGTCTGAGCAAGCTTGGCACTGC

10045 TCTGCTTCACTCTGCAGATTCGCAAGTCCGAGTADCTCTCAGAAAATTCCTGCACCTAT
TGTATGTATCTGCAAGATGAGAGGTATAAGCCACTGTCTAATGAAGTCCGCTTTCACAT
TAGAGCATGCCAAGCAAGCTCAATGTCAATTCAAAAGCAGAAAGAGTGCAGGTAGAGC
TATAGCTGAAGCTTATCTAGGGGAAGATTTGAGGGGAGCTCTANGCTCAACACAGCAACCA
CTTCCGCAAGAGCTCTCTCATGGTTTTCTCTCCGCAAGGCTCTATCTCTCAAGGAGAGAG
(C, A)

GGGTAACCCCTGTATATATGGCTCTTGTGTCAATGATCTGCTTTAATGGAACCTGCT
CTGTTTGAAGCAGATTATGTGTATAATTGAAAGCCGCGAGATCTTTAACTACGCCATT
[A, G, T]
ACCATATATGCACTTTTGCATGCTCCTCTCTCACTCCGCTGGGTGATATTTTCCCTTCC
TGTGTCCCTGTGTAAAGCAATATGGCTATTTACTCACTGTGATCTTTGGTTCTCTGGTGT
AGGTTGTCTTCCATTAGATCATATAAACAGGCCGCCAGCCAGCCCTCAATTAAGGCAAT
TTTGTGATCTGTGTGTGTGATGATTTGTGTCTTCACTCTCTCTGCCAGCATATGCGAG
AAGATTTGCAAGCATCAGAACACAGCCTGGAGTCTATGAGCAGATCAACTGATGTCT

ACTTACTGCTTTGTTTCAGTGTCACTATATAAATGCAAGTAAAAAGAACCTATCTCAT
ACATTTCTTGTACGATGAGTGGGTAAATATATAAAGCAATTAAGCAAGCTGCTGGCA
CTGAATAGATGTTAAATTAAGTATAGTATGTCAAACTGTCTTGTCTCCAGGAATTT
CTAAAGCAGACCAACATATGCAACCTTACACATACATATATGCAATACATGCAATAGAT
TTAATAAGAGCACTCAGAGAAGCAGTATATAACATTTAAGGCATAAATGGCATT
[T, C]
AAATAGCAGCACTTCCCAAGCTTTTCTGCAATCTTGCACACACAGAAATGTTAATGCTT
TTGTCTCTTATGGAATAAACAGGAATTTGGCGAAGCTATACAGAACTGTGTAA
AAGCAATCTTCTATTTTAAATATATACAAATATGATGAAAGCAAAATGCAAACTGT
TAGGCAAAATATTAATGTTAAATTTATCAAACTTAAACCTTTTCAATTTTTTTTT
TTTTTTTTTGTGATGAGATCTCTATCACTAGGCTGTAGCCGAGCTGCTGTATCTGAT

GCTAAGTCCCAATGCAATGGCGAGACAGAAATAAAGCAATGCACTAAATTTAACTGTAC
TTTGAAATGATGAGCAGCTTATCAATATTTGAGCAAGCAAAATCTGCAAGCTGTGT
CTAGAGAGGAGTTAGTAAAGCACTAAAGCAAGTATTAAGCAAGTTTGAAGATGTGAT
TGCAATCTGATTTATAGGCACTGATCAAAAAATTTGCATGGAATAAGATTAATTAAT
TGATGATAGGTCTCTGGAAATCTGGAGTTGAAGAGAAATTTCTAGGCCCTTTGATGCG
[G, T, A]
GGGCCCCCTTTGCAAGGCCCTTTTCTTATCTCAACCTTGGTTCTCTCTTATGCTTTGG
CAGAAATAGGTTATATACACATATTTGTTCACTGAAATTAATTTTAAACCCCTATTTAA

AGCTCTGATTTTTCODCTCAATCATATTGTGTTTGTATCTOCAMAGATTTATAAGT
GCATTTATTTTAAATAATTGTATATTGTACTTTCTAGCATGAAGTGTAGCAGCTTCTCA
GATATTGATGTACACTCTGAAGTCAGACATATTCCTGTTCGAGGACATGACACTTGGCA

18520 ATTTATOCATAAATTTGTCTGTCAATTGGTTTTCTAATCAATGGTGTGTGAAATGTCTTATTT
TCTTTATTTCACTTTGGCTCTGTATGCATTCGAATCAGACATTCGATTCCTGGGTGGGAC
TCAAGATTTAAACATAGGCTOCAGTGGAGCTCTCTCTCTGAGAGCTGATCATTAAG
CTGCATATTTAAGGCTCATTTTACACATCTCCAGGCGCTGTGCACCAATTTTATTCCT
CAGATTTGATTTTACACTTCAGACATAATATTCGATGATATATACTATAGTTAGTTTAC
[A, -, C]
AAATATGCACTGAGACATTTTAAATCTGAGACTTTTATTAAGTACAAATTTATTTGT
GGGCTGTCTTGGTGAAGTAAATGGTCTAATACAGGAGACAGAGAGAGCTOCAMATTT
GCAGTTCAGCATATCAGGCGAATGATGAGATATTTCTTGGTAAACAGAACATAGA
AGAGAGTAACTAGGCTTGGATGGAGCTGAGAGAGACTTCTTGCATTTAGCTGAGCT
CGAGATAGCTAGGATTTAGCGAGTGGCAACTGTCACTCTATTTCTGTAGACTTTAAG

18525 TCATAAATTTGTCTGTCAATTGGTTTTCTAATCAATGGTGTGTGAAATGTCTTATTTCTTT
ATTTCACTTTGGCTCTGATGCATTCGAATCAGGACTTGAATCCTGGGCTGGCACTTACA
ACTTAAACATAGGCTOCAGTGGAGCTTCTCTCTGAGAGAGCTGAAATGATTTAGCTGCA
TTATTTAAGGCTCATTTTACACATCTCCAGGCGCTGTGCACCAATTTTATTCCTCAGCA
TTGATTTTGAAGCTTCAGACATAATATTCGATGATATATACTATAGTTAGTTTAGCAAT
[-, T, A]
TGCAGTGGAGCAATTTTAAATCTGAGACTTTTATTAAGTACAAATTTATTTGGGGCT
TGTCTTGGTGGAGCAATTCCTCTAATACAGACAGGAGACAGAGCTOCAMATTTGAGAT
GTAGCATATCAGGCGAATGATAGAGATATTTGGTGAACAGAGGATAGAGACAT
GTAAGTAACTAGGCTTGGAGCTCAGGAGACTTCTTGCATTTAGCTGAGCTGAGG
ATAGTAGGAGTTAGCGAGTGGAACTGTCACTCTATCTTGGTACAGCTTAAAGCATAT

18525 TCATAAATTTGTCTGTCAATTGGTTTTCTAATCAATGGTGTGTGAAATGTCTTATTTCTTT

GTTCATTAACTCATACTATCATTTTCTTTTGTGGTAAGCAATTTAAATCTACCCCTCTTA
 GCAATTTTCAAGTATACAAATTTGTAGTAACCTCCAACTGACATATTTGACAAATGCATCTCC
 TAAACTCTTCCGCTCTCTGTGACTGAAATTTTGTATCTCTTGCAATACATCCCTGTAATC
 CCCCATTCTGCAAGCCCTGTGAACCACTGTTCCTACTCTGTCTCTTTGTAGTTTAAAT
 GTTTTAGATTTCCACATGTGAGATCAATGTGGAAATTTGTCTTTCTGTGCTGGCTTAATTTG

19325

TCAGAGATCTCATTTTATAAGGAAATAGAGGCCCTTTCTACCATAACTAAAGATTTAA
 TCTATATAGCACAAAAATACAATGTTGAGTAATCATTTTAAATTTATTTTAACTGACAA
 AAATTTGTCATATACATGTTATATATATATGCTATGTGTGATATATATATGATGTACAAAC
 ATGATATTTTGATATATCTATACACTGTGCAATGCAATATCTATCAATGAGCATGTTCA
 TTAACTCATACTTATCATTTTCTTTTGTGGTAAGCAATTTAAATCTACCCCTCTAGCAAT
 [G, T]

TTCAAGTATACAAATTTTGTAGTAACCTCAATCAGATATTTGACAAATGCTCTCTCTAACT
 TAGGCTCTTGTGTGACTGAAATTTGTATCTCTTACTGACATACCCCTGTAATCCCGCAT
 TCTCCACAGCCCTGTGTAACTGTTCCTACTCTCTGCTCTTTGTAGTTTAAATGTTTAA
 GATTTGCCACATGTGAGATCATGTGCAATTTGTCTTTTGTGCTGCTGCTATTTTCACTTAG
 CATAAATGTCAATCAAAATTCATCTCTTTGTGTCAAAATGACAGATATTTGTCTTTCTAT

19346

GAAATAGAGGCCCTTTCTACCATAACTAAAGATTTAAATCTATATAGCACAAAAATACAA
 TTTTGAGTAAATCAATTTTAAATTTATTTTAACTGACAAAAATTTGTCATATACATGTTA
 TATATATATCTATGTGTGTATATATATATGATGTACAGACATATTTTGTATATGTAT
 ATCACTGTGCAATGACATAATCTATCAATGCAATGTTCAATTAACCTAACATTCATCATTT
 TTTGTGTGTAAGCAATTTAAATCTACCCCTCTAGCAATTTTCAAGTATACAAATTTGTA
 [G, T]

TAACTCCAAATCAGATATTTGACAAATGCATCTCTTAACTTATGGCTCCTGTCTGACTGAA
 ATTTTGTATCTCTTGACTACATCCCTGTAATCCCGCAATCTCCACAGGCCCTGTGAAC
 CACTGTCTCTACTCTCTGCTCTCTTGAGTTTAAATGTTTAAATTTGACATGTGACATCAT
 GTGAAATTTGTCTCTTGTGCTGGCTATTTTCACTTAGCATATTTGACATGCAAAATTCAT
 CTCTGTTGTGATAAATGACAGATATTTGCTCTTGTGTGCTGCTAAATTTGTAGTCAATTTG

ATTTCACCTTGGCTCTGATGCAATGCAAACTGAGCACTGTAATGCTGGGCTGGCACTTACA
ACTTAACAATAGGGCTGAGAGCTGAGCTGCTCTTCTGAGAGCAAGTAATAGCTGCGA
TTTATAGGCTGATTTTAAAGCAATGCTCCGAGGCTTGTGCACTAATTTATCTCTGCA
TTGATTTTGAAGTTCAAGACATAATATGATGATATACTATAGTTAGTTTACCAAT
(- C, A)

TGCACTGAGGAAATTTTAAATCTGAGACTTTTATATGACTACAAATTTATGTTGGGCT
TGTCTTGGTGAGCTAATGCTCTAATACAGGAGAGAGGACAGACGCTTCCAAATTTGAGT
GTAGCATATACAGGGCAATGATAGACATATGCTGCTGATACACAAAGACATAGAGACCA
GTGACTTACCTTGGCTGATGGAGCTCAGGAGCACTGCTTGACATTTAGCTGCTGAGG
ATAGCTAGAGCTTATGAGCGAGCTGGAAGCTCTCATCTCATCTCTGAGCTTAAAGCAT

19789 CTGCTCAAGGCGACGAAAGACAGAAATGCTAGGACCAATTCAGACGACAGCCGATAAA
AAGACACCATATTTCTATGCAAAAGGTCAACTCAATTCGAACCATTTGTAAAAACCAATTTG
TATATATAAAGTATATACGAGATCTCATTTTATAGGAAATAGAGAGGCCCTTCTCTACCA
TAAAGCTAAGAGTATTAATCTATATGACCAAAAATACAACTGTGCAATTAATCATTTTAAATTT
ATTTTAACTGTCAAAAAATTTGTGCATATACAGTGTATATATATATATGTATGTGTATATAT
(T, C, A)
TATATGATGTACAGACATATATTTTGTATATATGTATACACTGTGGAACTCAATTAATCTAT
CAATGGACATGTTCTAATACCTCATCTTATCATTTTTTTTGTGTAGGACATTTAAATCT
TAGCCCTTTAGCAATTTTTCAGTATACAAATTTGTTAGTAACTGCATACAACTTGCAT
ATGCATCTCTCAAACTATAGCTCTCTGTCTGCAAGAAATTTGTATCTCTTCACTAAGCAT
CCCTCAATGTCCGCAATTTGTCCGACAGGCCGCTGTGAAGCACTGTCTCACTGTCTGTCTT

19259 TTTCATATGCAAAAGTCACTCAATTGAAACATTTGTAAACCAAMTTTGACATATAAA
AGTATACAGAGATCATATTTTATAGGAAATAGAACGCTTTTGTACCATAAACATAAAG
ATTATATCTATATAGCAGAAAATACAAATTTTGTATTTTATTTTATTTTAAAC
TGACAAAAATTTTGACATACATGTATATATATATATATTTGTGTATATATATATATG
TACAAACATGATTTTGTATATATATATACACTGGAAATGCAATCTATCAATGCAAC
[C, T]

20845 TGTATTCTGGAACCTTTGTAGATCAGTTGCACATAAATGTGTGGGTATATTTCTGGAAGCTCT
TATATCTGTTTATTAAGTTAATAAGTCTCTTTTATTAAGAAGCTCTATCTGTGTTTGGTGA
CTAGAGAGCTGTAGTCAAATTTAGATCAGGTAGTATGATGCATCTCAGCTTTGCTCTTTT
TGTCAAATTTGCTTTGGCTATAGCTTTTTTATTCATCAAGCAATATAGGCTCTTTT
TTTTTTTTTGCAATCTGTGAATAATGCAATTCGAATTTTGAATGCAATTCGAATGAATCT
[-, ?]
TGGTAGTATGCAATTTTAAACAGTATTAATGCTTCAATTAATGAACAGGCTATTTT
GCAATTTGTGTTTCTTCAATTTCTTTCACCAGTGTTTTTCTTAATTTAATGTGTTTA
TTTCTATAGGTTTGTGATACAGCTGTGTTTGTGTTATAGTAGTCTTTTACTGTGAT
TTGTGAGATTTTGAATGCCACCAATCACTAAGCAGATACAGCTATGCCAAATTTGTAGTCT
TGATATCCCTCACTGCTCCCTGCAACCAATTTCCCGCAATGCGCAAGATGCATGATCAATCT

20845 TGTTACTGCACCTTTGTAGATCAGTGCACATAAATGTGTGGTCTATTCTGCAGCTCT
TTATCTGCTTTTATTAGTTATATATGCTCTTTTATAGAGAGCTATATGCTGTTTGTGTG
CTAGAGCTGTGTATGCATTTTCCAGATCAGGTACTATATGCATGCACAGCTTTGCTCTTTT
TGCTCAAAATTCGCTTGCGTATTCAGTTTATTCATGCATACGAATTTTAGGGCTTTT
TTTTTTTTGCATTACTGTCAATAATGCCATTGCCAATTTTATGCGATTGCATTGAATGCT
[7, C]
TGGCTACTATGCATATTTTACAGTATTAATGCTTCCAAATTAATCAACAGGCTATTTT
GCAATTTGTTGTTTCTTCAATTTCTTTCACCACTGTTTCTTCTAATTAAATGTGTTTA
TTTGCATAGGCTTTGGTACAGAGGTGTTTGGTATATGATAGTCTTCTTACTGTGCTA
TTGTGAGATTTTTCATGCACCACTCACTAAGCAGTATACACTATACCCAAATTTGTAGTGT
TGATATGCCATTCCTGCGACCAATTCGCCCAACGCGCAAGTCCCAATGTATCATCTC

22234 AGAAACTTTTGTAGTTTAATTAAGTCCACCTATTTATCTTTTGGTGTGTGTCTTTTTC
GGGTGTCTTTTGTGGCTTGGTTTGCATCTGCTTTTGGTCTTGGTCATCAAGCTTT
TGCCTAAGCCAAATATCTAGAGGGCTTTTCTGATGTTCTAGAAATTTTATGCTTCAGTC
TTAGATTAAAGTCTCTCATCATCTTCACTTGAATTTTGTATAAGGTGAGAGATCAGAT

CCAGTTTCATGCTTACATGCTTCCAAATATCCAGTACAATTTGTAATAGG
(1, C)
TAATATTAAGCTTATATATTAAGGTTGCTATTTGGGTACATATTTATTAAC
TATCATATGCTGATGGAATGACCTTTCTCATATATATGCTGCTGCTGCT
TTTACAGTTTGTCTTAAAGCTTAAATTTGCTGATAAAGTTGAGTACCTTTGCTG
TTTGGTTTCTATTTGCAATGCAATATTTTCCAAAGCTTGGATTCAGTCTATGCTG
TTCTTAAGATGAATGAGATGCTGATGCTTGGGCTATGCTTGGTCTGTTTATTCATC

22234 AGAACTTTTATGTTAAATTAAGTCCACCTATTTATCTTTGCTTCTGTTGTTT
GGGTTGTTTCTTTGGCTTGGTTTTCATCTGCTTTGGCTTCTGCTCATGAGTCT
TGGCTAAGCCAAATATCTAGAGGCTTTTCTGATGCTTACAAATTTTATGCTT
TTAGATTAAGTCTTGAATCTTGAATTTTGTATAGGTGAGATGAGAT
CCAGTTTATGCTTCTACATGCTTGGCAATATCCAGTACAATTTGTAATAGG
(6, T)
TAATATTAAGCTTATATATTAAGGTTGCTATTTGGGTACATATTTATTAAC
TATCATATGCTGATGGAATGACCTTTCTCATATATATGCTTCTGCTGCT
TTTACAGTTTGTCTTAAAGCTTAAATTTGCTGATAAAGTTGAGTACCTTTGCTG
TTTGGTTTCTATTTGCAATGCAATATTTTCCAAAGCTTGGATTCAGTCTATGCTG
TTCTTAAGATGAATGAGATGCTGATGCTTGGGCTATGCTTGGTCTGTTTATTCATC

22247 TTTAATTAAGTCCACCTATTTATCTTTGCTTCTGTTGTTTCTTGGGTTGTTT
TTGGCTTGGTTTTCATCTGCTTTGGGTTCTGCTCATGAGTCTTGGCTAAGCCAA
ATCTAGAGGCTTTTCTGATGCTTACAAATTTTATGCTTGAATTTAGATTAAG
CTTGAATCATCTTGAATTTTGTATAGGTGAGATGAGATCCAGTTTGAATGCT
TCTACATGCTGCTTCCAAATATCCAGTACAATTTGTAATAGGCTTAAATTTAAG
(6, T)
TTTATATATTAAGGTTGCTATTTGGGTACATATTTATTAAGTATCATATGCTG
TGAATGATGACCTTTCTCATATATATGCTTCTGCTGCTTTTACAGTTTGG
TCTTAAGGCTAATTTGCTGATAAAGTTGAGTACCTTTGCTGCTTTTGGTTCTAT

CACTTTACTCTACCAAGTCCCTCCATTTATGCTTTGATGCTATAATTAAGT
TTGGAGATGCTGCTTCTATGCTATGCTTAAACAAATTAATGAGTACAGTATTT
TAATAGTTTGGCTTTAATCTTATAGTATGAAATTAATTAACATACCACTAC
(-, A)
TTAATAGGCTATCTAAATGACTATGATTTACCTTTATGAGTATGATTTTGTCT
ATTTTCAATGCTTATTAATGATTTCTTCAATTTCACTTGGAGATTCAGATTAAG
TTTGTAGATGCTTCTAGTATGCTTGAAGCTTCACTTTTGTATCTGAGATGCT
TTAGCTGCTGCTTCAATTTGAATATAGCTTTTGTGATGATGAATGAGCAAAATG
TTTTTAATTAAGCAAGTCCAGGTAAGCAGAAATGCTTTTGTGTTTCTGAGAT

23421 CTTCATTTCACTTGGAGAAATGCAATAGCAATTTTGTAGATGCTTATGATG
TGACAGCTTCACTTTTGTATCTGAGATGCTTTACCTTCTGCTTCAATTTGAAT
TAAGTTTGTCTCATGATGAATGAGCAAAATGTTTGTATTAATGCAAGTCCAG
GCTAAGCAGAAATGCTTTTGTGTTTCTGAGAGCTTCACTTCTTGGCCAG
GCTGAGTCCAGTGGGCAATGCTGAGCTTADGCAAGCTTGGCTCCAGCTTCAAG
(A, G)
ATTTCTTGGCTCAGCTTCTGAGTATGCTGCAATACAGGATGCAAGCAATGCTG
CTAATTTGCAATTTTATGAGAGGCTTCTGCTATGCTGCTGCTGCTTGAAG
ADGCAAGCTGAGATGATGCTGCTTGGCTTCCAAAGTCTGAGATGAGGCTGCA
GGCACTGCTGCTGCTGAGATGCTTATTTATGCTGAGTGAAGAGAAATGCA
AAATTAATTTCACTTACCTAATGCTGCAAGCTTCAATTAAGTAAATGCAAA

25582 CCAAGGCAATAGGAGTTGAGCAAGCTAAGGCAAGCTAAGCTATGCTTCTGAAAT
CAGGCAAGGCTCTTCTGATGCTGAGATCATTTCTAAGATAATCTTTGCTGAG
TTTGAAGGAGCTGAAGCTTCAATGAGAAATAGCAGCACTTTATCTGAAGCAAA
GGGCAATTTGGCTCATGATCATATATGCTTAAATATAGCAATTTAATGAT
TAATATAGGCTTCAAGGCAATATCTTATTTTGGCTTGAATGCAATGTAAGCA
(1, C)
GCTTATAGCAATTTACAGTCTGAGCACTTTAAGTATCTTCAATGCTGCTGCAAA

TTGATGCAATATTTTTCAGGCTTGGATTCAGTATGCTGCTTCAAGATGA
AATGAGATGCTTGAAGGCAATGCTTGGCTTCTGTTTATTCATTCATTCAGGCT
22234 GTTCTGCTGATGAAGTCTTGGCTAAGGCAATATCTAGAGGCTTTTCTGATGCT
CAATTTTATGCTTGAAGTCTTGAATTAAGTCTTGAATGCTTGAATTTTGT
ATAGGTGAGATGAGATGCAATGCAATTTGCTTCAATGCTTGAATTTGCTTGA
GTACAAATTTGTAATAGGCTTAAATTTAAGGCTTTATATTTAGGCTTCTATTT
GGTACATATTTATTAAGTATGATATGCTTGAATGCAATTTGCTTGAATTT
(A, G)
TAATGCTTCTGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
GTTGAGTATGCTTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
TGGATTCAGTCTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
GGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
ATTCAGGCTAATTTATGAGAGAGAGGCTTCTTCTTCTTCTTCTTCTTCTTCTT

23033 ATCTTTTGTGCTCTACTATAGGTTTGTCTTGTGCTTACATGAGGCTACATAGG
ATAGTATTAAGGCTATTTTAACTGATAGGCTTAACTTCAAGCTTAAAAAAT
ATAGCTTTTACTCTAGCAAGTCCCTCCATTTATGCTTCTTGAATGATTAATTA
GTTTGAAGATGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
TTTTAATGTTTGTGCTTTAACTTTATGCTAGATGAGATTAATTAAGTATGCT
(T, -)
ACATTAATAGGCTATTTAAATGCTATGATTTAGCTTTATGAGTGAATTTGTT
TCAATTTCTGCTTCTTAAATGCTATTTCTTCTTCTTCTTCTTCTTCTTCTTCT
TTTTTGAAGTGGCTCTAGTGGCTGAGGCTGAGGCTTCTTCTTCTTCTTCTTCT
TCTTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
TCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT

23035 TTTTGTGCTCTACTATAGGTTTGTCTTGTGCTTACATGAGGCTACATAGGCA
GTTATTAAGGCTATTTTAACTGATAGGCTTAACTTCAAGCTTAAAAAATGATA

TCCAGAGGAGGCTTGAATGATGATTAAGTATGCTGAGTGAAGTGAAGTGT
GGGAGAGGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGAT
CTTATAGGAGGAGGCTTGAATGATGATGATGATGATGATGATGATGATGAT
ATGCTTGAAGAGGAGGCTTGAATGATGATGATGATGATGATGATGATGATGAT
28407 CCTCTGAGAGAGGCTTGTATGATGATGATGATGATGATGATGATGATGAT
TGCTGAGGCTTGTATGATGATGATGATGATGATGATGATGATGATGATGAT
TGGATTAAGGAGGAGGCTTGTATGATGATGATGATGATGATGATGATGATGAT
TGCTGAGGCTTGTATGATGATGATGATGATGATGATGATGATGATGATGAT
TCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
TCCAGAGGCTTGTATGATGATGATGATGATGATGATGATGATGATGATGAT
(C, A)
CTAATTTGTCTGAGGAGGCTTGAATGATGATGATGATGATGATGATGATGAT
GCTGAGGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGAT
GAGTCTGAGGCTTGAATGATGATGATGATGATGATGATGATGATGATGAT
AGGATGAGGCTTGAATGATGATGATGATGATGATGATGATGATGATGAT
GGAGGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT
28473 CAAGCTTGTCTGAGGCTTGAATGATGATGATGATGATGATGATGATGAT
TACAGGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGAT
TGTGAGGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGAT
GGGCTGAGGCTTGAATGATGATGATGATGATGATGATGATGATGATGAT
TTGCTGAGGCTTGAATGATGATGATGATGATGATGATGATGATGATGAT
(C, T)
GCAATATATGCAATATTTTGTGAGGAGGCTTGAATGATGATGATGATGAT
GAGGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT
AGGAGGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGAT
TCACTGAGGCTTGAATGATGATGATGATGATGATGATGATGATGATGAT
CTGATGAGGCTTGAATGATGATGATGATGATGATGATGATGATGATGAT

25844 TGGCAACATAGGAGACCCCTGCTTTACAAAAAAGAGAGAGATAGCCAGGAT
GCTGTTGCACTGCTTGTATTCCTGCTACTTGGGGAETGAGGAGGAGATCACTTGAGC
TCAGAGTTCAAGGTTAAGGTCAGCAATGTTCAAGGCACTGCTTCCAGGCTGATTCACA
GGCAGAGCCCTGACTCTAAACAAAAAAGAAAAACAAATATTTAAGTAAATTCACAAACATA
GCAGAAAAATATAGCATGCTTTATCACTTTGATATGACACCAACAGCTACTTAAATAGA
(S, A)
TCATCAATTCAGTAAATTTGTTGCTGCAAGCTAAGGTCGCAACCCAGGCGCATCTTCT
TAGGTCCTCTCACTGCTGCTCATGAGTACAGCAGCAGAGCAATTCAGCAGCTAGCTC
TTCCCTTCAAGAACAAAGCTCTTGTAAAGGCACTAGGCGCAGCACTTCTTCTTCTC
CTGCACTCTCTTTAATTCCTCTCTTCTTAAAGGATCAGGCTGCTTCTAGTATTTGGG
TCTTCACACAGCCCTGCTGCTGCAAAAACCAAGGATATGATTCCTCTCTTGTACATAA

25384 CTTCAGGGAAGGATAGATCTTGGTCCCTATTTGAGCCCAAGGATCAGTTAGTTTTAC
AAGGACAATGATATTCCTGCTGACATGCTTTTGGCCATGCTTCAAAAGCAGTCCACACA
ATGTAAGCTACTGCTCATAGGCTCAATGAGTCCAGCTTCAAGCAGAGAAATATTTTC
ATGAGTAAGTCCAGTCCGCTCTGTTATAGGGAAGGCAATGTTGGAGCTCCAGCT
CAAAATTCAGAGTGAACAAATTAAGTCTAAAGTTCAAAAGTTTCAATGGCATTTGCTGG
(A, -)
AAAAATATCACTTTACTGCTACTTCAGACTCTTGTACTAGTATTTTACTATAGTCAGA
AGAAACATCAATTTTCAAGTATCACTTTCTTCCCTTGTCTTCAGAACTGCAATTCG
GCAGAGTTTGGCATGATTCAGTTAAGGTAACCAATGGCTTGAATCTGCTCCACTTCAG
AGTCACTCCAGAGCCCAAGGCTCTTACTTTCCCAAGCAATTTATCTCAGAGCCAA
CAATGGGATGATTTTGCAGCTGAGAACTCTCTGAATGTTAGATTCAGGATACATGA

28417 GAGCCCAAGGATCAGTTAGTTTTACAAAGGACAATGATATTCCTGCTGACATGCTTTT
TGGCATGCTTCAAAAGCAGTCCACAAATGTAAGTACTGCTCATAGGCTCAATGCACTC
CACTTCAAGCAGAGAAATATTTCACTAGTAACTCCAACTGCGCTTGTATAGGG
AAGGCAATCATGTTGAGGCTCCAGCTCAAAATCTCAGAGTGAACAAATTAAGTCTAAG
TTCAAAAGTTCAATGGCATTTGGTGA AAAATATCACTTTACTGCTACTTCAGACTT

(A, C)
TTGTAAGTATTTTACTATAGTCAGAAAGACATCTTTTCAAGTATCACTTTCTTT
CCCTCTGCTCTTCAGGAATGCAATGGCAGGAGTTTCCATGATTCAGTTAAGGTAAC
CAATTCCTTGAATTCCTGCTGCACTTCAGAGTACTCCAGAGCCAGGCTCTTACTTT
CCCAAGCAATTTATCTCAGGCGCAAGATGGGATGATTTGCACTGAGAACTCTC
TGAATGTTAGATCTCAGGATACAATGATTAAGCTACTTTGTTTTGCAAGTTAAATTA

29265 TATGCAAGTAAATAGTGCATGATGCTCACTGTCAAAAATTCACACTAGAAATCAT
GTAGAAATAAAATTTTAAATCTCACTTCACTTAGGCGACATTCATGCTGACCAATCC
TACTGCTTTTCTTAAAAACAGAAATATTTGCTGCTCAATTTTCAGACTTTTCTGATAC
ATTTTATATGTAGAAATGTAGCAATGATTTGTTATAGATGCTCATTCCTATATTTGTA
TTGATTTTTCCTTAATAAAATTCAGCTTATTCCTTATCATTCCTTTATGCTATTCT
(A, G)
TAATATCAATGCTATATAATTAATTAAGTATTTTCTTATTTGGGCAITTAAGTTATTTTC
TAGTTTTAAAAACATGCTTGTCAATGGCAGCAAGGCAAAATTCAGAAATGGGATCTAA
TTAAACTAAGAGGCTTCTGCACAGCAAAACAACTACCATCAGCTGAATGGCAGGCTA
CAGAAATGGGCAAAATTTTTCAGCTCTCATCTGCAAGGCTAATATCCAGAACTCT
ACATGCACTCAACAAATGTACAGAAAAAACAGCCCATCAAAAGTGGTGAAGGA

29484 GTCATCAITTCCTATATGTTATTCATTTTTTCACTTAATAAAATTCAGCTTATTCCTT
ATCATTCCTTTATGCTATTCCTGTAATGAAATGCTACTATAATTTATTTAACTATTTCT
TATTTGGCAITTAAGTTATTTCTAGTTTTAAAAACATGCTTGTCAATGGCAGCAAGGCT
AAATTCAGAAATGGGATCTAATTAAGTAAAGGCTTCTGCAGCAAAACAACTACC
ATCAGACTCAATGGGAGGCTACAGAAATGGGAGAAATTTTTCAGGCTACTCATCTGAC
(A, G)
AAGGCTAATATCCAGAACTACAAATGCACTCAACAAATGTACAGAAAAACAAAGCC
CATCAAAAGTGGGTGAAGGATATGACAGACACTTCTCAAAAGAGACATTTAGGAGC
CAAAAGACACATGAAAAATGCTATGCTCACTGGCATCAGAGAAATGCAATCAAAAC
CACAAATGATACCATCTCAGAGGATAGAAATGGCAATCAATTAAGTGAAGAAACAA

CAGTCTGAGAGGATGAGAAATAGGAGACTTTTACACTGTTGCTGGCAGAGAA

30417 ATTCACTACCTAAAATCTATATATAAAAAATCCCTCCCTGAAATTCAGATCCTTGA
GACAAACACCCAGGCTCAAAACAAATTTGTTAACACTGGACAGTCTGCTGCTGAC
TTTCCATTTTGTCACTATTTTGTGAGTGTATACCAATATCCAGCCAGTTAAACAAATAT
TTCTGCTTTTCTGCTGCAAAACCAAAATAAATTAACAAATCAATAAAGCTAAATTT
CTAAATAACTCACTTTCTCTATATATCTCTTCTTCTGCAAAATGGGTTAGGTTAGT
(T, -)
CTTTAAAGCATGATGATAAATTTGCTGAAATACAAATTCAGGCTGAGCATACTAGG
TATAATTTCTGCTGCTGAGGCTCTTACTTATTTGGGTCAAAATAAACAGTTTATTA
AGCTTATTAATATTCAAATTCATTAATCTCTTAAACAAATATGTTCCCTGAGTTTCAAT
TGGCAATATTTATTTCTCAGGTTGGCAGGCTCTTCAAACTCTGCTATTTTTCATA
TCCAAATTTACTTTAAATATTTTAGAAAGAGGCTCTGTTAAATTTCTTAATATTAATTA

30783 TTTTCTGCTCTCTGGGCTTACCTATTTGGGCTCAAAATAACAGTTTATTAAGCTT
ATTAATATTCATTTCAATATCTCTTAAACAAATATGTTCCCTGGTGGTTTCAATGGCA
ATTAATTAATTTGAGGTTGGCAGGCTCTTAAACTCTGCTGATTTTTCATATCCAA
TTTTACTTTAAATATTTTAGAAAGAGGCTCTGTTAAATTTCTTAATATTAATTAATTA
TTGTTTTTCACTGACATTTTGTCAATTCAAAACCTTAAAAATATGAATCATTTTTTC
(C, G)
AAATATGTCACAGACATTTTGTAAATAGAGACAGAAACAGGCAATTAAGAGAG
ATAAATATTCATATACCTTATATTTCTGTCAGACATTTTATACCACTGTCGCAAAA
TTGATATCATATAAATGATAACAGTTTCAAGAGGCAATTCCTTATCCCTTAACTCTCA
AATTAGAACTTTCAATAGGAGAGGAGGGAAGCATATATCCCTTTCAAGGTCGAA
GAAATGCTATTTGGCATTCAGCATGCTCTTCAAGCTTAAAAAATGCACTGCAAAA

FEATURES:
3 1/2" TUB: 1-105
Steel Cordon: 100
Pump Cordon: 1720
3 1/2" TUB: 1720-3177

WO 02/34922 A2



SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA,
ZW.

(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD,
TG).

Published:

— without international search report and to be republished
upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.

WO 02/34922

PCT/US01/42528

ISOLATED HUMAN DRUG-METABOLIZING PROTEINS, NUCLEIC ACID
MOLECULES ENCODING HUMAN DRUG-METABOLIZING PROTEINS,
AND USES THEREOF

5 RELATED APPLICATIONS

The present application claims priority to provisional application U.S. Serial No. 60/241,745, filed October 20, 2000 (Atty. Docket CL000897-PROV), and is a continuation-in-part of application U.S. Serial No. 09/739,456, filed December 19, 2000 (Atty. Docket CL000897) continuation-in-part of 09/818,647 filed, March 28, 2001 (Atty. Docket CL000897-CIP) and U.S. Serial No. 09/852,067 filed, May 10, 2001 (Atty. Docket CL000897-CIP-B).

10 FIELD OF THE INVENTION

The present invention is in the field of drug-metabolizing proteins that are related to the omega-hydroxylase cytochrome P450 drug-metabolizing enzyme subfamily, recombinant DNA molecules and protein production. The present invention specifically provides novel drug-metabolizing peptides and proteins and nucleic acid molecules encoding such protein molecules, for use in the development of human therapeutics and human therapeutic development.

BACKGROUND OF THE INVENTION

Drug-Metabolizing Proteins

20 Induction of drug-metabolizing enzymes ("DMEs") is a common biological response to xenobiotics, the mechanisms and consequences of which are important in academic, industrial, and regulatory areas of pharmacology and toxicology.

For most drugs, drug-metabolizing enzymes determine how long and how much of a drug remains in the body. Thus, developers of drugs recognize the importance of characterizing a drug candidate's interaction with these enzymes. For example, polymorphisms of the drug-metabolizing enzyme CYP2D6, a member of the cytochrome p450 ("CYP") superfamily, yield phenotypes of slow or ultra-rapid metabolizers of a wide spectrum of drugs including antidepressants, antipsychotics, beta-blockers, and antiarrhythmics. Such abnormal rates of drug metabolism can lead to drug ineffectiveness or to systemic accumulation and toxicity.

WO 02/34922

PCT/US01/42528

For pharmaceutical scientists developing a candidate drug, it is important know as early as possible in the design phase which enzymes metabolize the drug candidate and the speed with which they do it. Historically, the enzymes on a drug's metabolic pathway were determined through metabolism studies in animals, but this approach has now been largely supplanted by the use of human tissues or cloned drug-metabolizing enzymes to provide insights into the specific role of individual forms of these enzymes. Using these tools, the qualitative and quantitative fate of a drug candidate can be predicted prior to its first administration to humans. As a consequence, the selection and optimization of desirable characteristics of metabolism are possible early in the development process, thus avoiding unanticipated toxicity problems and associated costs subsequent to the drug's clinical investigation. Moreover, the effect of one drug on another's disposition can be inferred.

Known drug-metabolizing enzymes include the cytochrome p450 ("CYP") superfamily, N-acetyl transferases ("NAT"), UDP-glucuronosyl transferases ("UGT"), methyl transferases, alcohol dehydrogenase ("ADH"), aldehyde dehydrogenase ("ALDH"), dihydropyrimidine dehydrogenase ("DPD"), NADPH:quinone oxidoreductase ("NQO" or "DT diaphorase"), catechol O-methyltransferase ("COMT"), glutathione S-transferase ("GST"), histamine methyltransferase ("HMT"), sulfotransferases ("ST"), thiopurine methyltransferase ("TPMT"), and epoxide hydroxylase. Drug-metabolizing enzymes are generally classified into two phases according to their metabolic function. Phase I enzymes catalyze modification of functional groups, and phase II enzymes catalyze conjugation with endogenous substituents. These classifications should not be construed as exclusive nor exhaustive, as other mechanisms of drug metabolism have been discovered. For example, the use of active transport mechanisms been characterized as part of the process of detoxification.

Phase I reactions include catabolic processes such as deamination of aminases, hydrolysis of esters and amides, conjugation reactions with, for example, glycine or sulfate, oxidation by the cytochrome p450 oxidation/reduction enzyme system and degradation in the fatty acid pathway. Hydrolysis reactions occur mainly in the liver and plasma by a variety of non-specific hydrolases and esterases. Both deaminases and amidases, also localized in the liver and serum, carry out a large part of the catabolic process. Reduction reactions occur mainly intracellularly in the endoplasmic reticulum.

Phase II enzymes detoxify toxic substances by catalyzing their conjugation with water-soluble substances, thus increasing toxins' solubility in water and increasing their rate of excretion. Additionally, conjugation reduces the toxins' biological reactivity. Examples of

WO 02/34922

PCT/US01/42528

phase II enzymes include glutathione S-transferases and UDP-glucuronosyl transferases, which catalyze conjugation to glutathione and glucuronic acid, respectively. Transferases perform conjugation reactions mainly in the kidneys and liver.

5 The liver is the primary site of elimination of most drugs, including psychoactive drugs, and contains a plurality of both phase I and phase II enzymes that oxidize or conjugate drugs, respectively.

Physicians currently prescribe drugs and their dosages based on a population average and fail to take genetic variability into account. The variability between individuals in drug metabolism is usually due to both genetic and environmental factors, in particular, how the drug-metabolizing enzymes are controlled. With certain enzymes, the genetic component
10 predominates and variability is associated with variants of the normal, wild-type enzyme.

Most drug-metabolizing enzymes exhibit clinically relevant genetic polymorphisms. Essentially all of the major human enzymes responsible for modification of functional groups or conjugation with endogenous substrates exhibit common polymorphisms at the genomic level.
15 For example, polymorphisms expressing a non-functioning variant enzyme results in a sub-group of patients in the population who are more prone to the concentration-dependent effects of a drug. This sub-group of patients may show toxic side effects to a dose of drug that is otherwise without side effects in the general population. Recent development in genotyping allows identification of affected individuals. As a result, their atypical metabolism and likely response
20 to a drug metabolized by the affected enzyme can be understood and predicted, thus permitting the physician to adjust the dose of drug they receive to achieve improved therapy.

A similar approach is also becoming important in identifying risk factors associated with the development of various cancers. This is because the enzymes involved in drug metabolism are also responsible for the activation and detoxification of chemical carcinogens. Specifically,
25 the development of cancer is associated with alterations in the expression of these enzymes.

WO 02/34922

PCT/US01/42528

Abnormal activity of drug-metabolizing enzymes has been implicated in a range of human diseases, including cancer, Parkinson's disease, myotonic dystrophy, and developmental defects.

5 Cytochrome p450

An example of a phase I drug-metabolizing enzyme is the cytochrome p450 ("CYP") superfamily, the members of which comprise the major drug-metabolizing enzymes expressed in the liver. The CYP superfamily comprises heme proteins which catalyze the oxidation and dehydrogenation of a number of endogenous and exogenous lipophilic compounds. The CYP superfamily has immense diversity in its functions, with hundreds of isoforms in many species catalyzing many types of chemical reactions. The CYP superfamily comprises at least 30 related enzymes, which are divided into different families according to their amino acid homology. Examples of CYP families include CYP families 1, 2, 3 and 4, which comprise endoplasmic reticulum proteins responsible for the metabolism of drugs and other xenobiotics.

15 Approximately 10-15 individual gene products within these four families metabolize thousands of structurally diverse compounds. It is estimated that collectively the enzymes in the CYP superfamily participate in the metabolism of greater than 80% of all available drugs used in humans. For example, the CYP 1A subfamily comprises CYP 1A2, which metabolizes several widely used drugs, including acetaminophen, amitriptyline, caffeine, clozapine, haloperidol, imipramine, olanzapine, ondansetron, phenacetin, propafenone, propranolol, tacrine, theophylline, verapamil. In addition, CYP enzymes play additional roles in the metabolism of some endogenous substrates including prostaglandins and steroids.

Some CYP enzymes exist in a polymorphic form, meaning that a small percentage of the population possesses mutant genes that alter the activity of the enzyme, usually by diminishing or abolishing activity. For example, a genetic polymorphism has been well characterized with the CYP 2C19 and CYP 2D6 genes. Substrates of CYP 2C19 include clomipramine, diazepam, imipramine, mephenytoin, moclobemide, omeprazole, phenytoin, propranolol, and tolbutamide. Substrates of CYP 2D6 include alprenolol, amitriptyline, chlorpheniramine, clomipramine, codeine, desipramine, dextromethorphan, encainide, fluoxetine, haloperidol, imipramine, indoramin, metoprolol, nortriptyline, ondansetron, oxycodone, paroxetine, propranolol, and propafenone. Polymorphic variants of these genes metabolize these substrates at different rates, which can effect a patient's effective therapeutic dosage.

25

30

WO 02/34922

PCT/US01/42528

While the substrate specificity of CYPs must be very broad to accommodate the metabolism of all of these compounds, each individual CYP gene product has a narrower substrate specificity defined by its binding and catalytic sites. Drug metabolism can thereby be regulated by changes in the amount or activity of specific CYP gene products. Methods of CYP regulation include genetic differences in the expression of CYP gene products (i.e., genetic polymorphisms), inhibition of CYP metabolism by other xenobiotics that also bind to the CYP, and induction of certain CYPs by the drug itself or other xenobiotics. Inhibition and induction of CYPs is one of the most common mechanisms of adverse drug interactions. For example, the CYP3A subfamily is involved in clinically significant drug interactions involving non-sedating antihistamines and cisapride that may result in cardiac dysrhythmias. In another example, CYP3A4 and CYP1A2 enzymes are involved in drug interactions involving theophylline. In yet another example, CYP2D6 is responsible for the metabolism of many psychotherapeutic agents. Additionally, CYP enzymes metabolize the protease inhibitors used to treat patients infected with the human immunodeficiency virus. By understanding the unique functions and characteristics of these enzymes, physicians may better anticipate and manage drug interactions and may predict or explain an individual's response to a particular therapeutic regimen.

Examples of reactions catalyzed by the CYP superfamily include peroxidative reactions utilizing peroxides as oxygen donors in hydroxylation reactions, as substrates for reductive beta-scission, and as peroxyhemiacetal intermediates in the cleavage of aldehydes to formate and alkenes. Lipid hydroperoxides undergo reductive beta-cleavage to give hydrocarbons and aldehydic acids. One of these products, trans-4-hydroxynonenal, inactivates CYP, particularly alcohol-inducible 2E1, in what may be a negative regulatory process. Although a CYP iron-oxene species is believed to be the oxygen donor in most hydroxylation reactions, an iron-peroxy species is apparently involved in the deformylation of many aldehydes with desaturation of the remaining structure, as in aromatization reactions.

Examples of drugs with oxidative metabolism associated with CYP enzymes include acetaminophen, alfentanil, alprazolam, alprenolol, amiodarone, amitriptyline, astemizole, buspirone, caffeine, carbamazepine, chlorpheniramine, cisapride, clomipramine, clomipramine, clozapine, codeine, colchicine, cortisol, cyclophosphamide, cyclosporine, dapsone, desipramine, dextromethorphan, diazepam, diclofenac, diltiazem, encainide, erythromycin, estradiol, felodipine, fluoxetine, fluvastatin, haloperidol, ibuprofen, imipramine, indinavir, indomethacin, indoramin, irbesartan, lidocaine, losartan, macrolide antibiotics, mephenytoin, methadone, metoprolol, mexilitene, midazolam, moclobemide, naproxen, nefazodone, nicardipine,

WO 02/34922

PCT/US01/42528

nifedipine, nitrendipine, nortriptyline, olanzapine, omeprazole, ondansetron, oxycodone, paclitaxel, paroxetine, phenacetin, phenytoin, piroxicam, progesterone, propafenone, propranolol, quinidine, ritonavir, saquinavir, sertraline, sildenafil, S-warfarin, tacrine, tamoxifen, tenoxicam, terfenadine, testosterone, theophylline, timolol, tolbutamide, triazolam, verapamil, and vinblastine.

Abnormal activity of phase I enzymes has been implicated in a range of human diseases. For example, enhanced CYP2D6 activity has been related to malignancies of the bladder, liver, pharynx, stomach and lungs, whereas decreased CYP2D activity has been linked to an increased risk of Parkinson's disease. Other syndromes and developmental defects associated with deficiencies in the CYP superfamily include cerebrotendinous xanthomatosis, adrenal hyperplasia, gynecomastia, and myotonic dystrophy.

Omega-Hydroxylase Cytochrome P450

The novel human protein, and encoding gene, provided by the present invention is related to the omega-hydroxylase cytochrome P450 family, which includes, for example, cytochrome P450 4A4 (CYP4A4), cytochrome P-450p-2, prostaglandin omega-hydroxylase, and laurate omega-hydroxylase. Omega-Hydroxylase Cytochrome P450 proteins catalyze omega- (including omega-1) hydroxylation of prostaglandin A and fatty acids such as caprate, laurate, myristate, and palmitate (Yoshimura *et al.*, *J Biochem* (Tokyo) 1990 Oct;108(4):544-8). CYP4A4 is elevated during pregnancy (Palmer *et al.*, *Arch Biochem Biophys* 1993 Feb 1;300(2):670-6).

Matsubara *et al.*, *J Biol Chem* 1987 Sep 25;262(27):13366-71; Yamamoto *et al.*, (1984) *J Biochem.* (Tokyo) 96, 593-603; Yokotani *et al.*, *Eur J Biochem* 1991 Mar 28;196(3):531-6; and Johnson *et al.*, *Biochemistry* 1990 Jan 30;29(4):873-9.

Cytochromes, such as the protein provided by the present invention, have many utilities, in addition to those described above. Cytochromes not only metabolize normal physiological substrates but also neutralize environmental toxins. In addition to oxidizing steroids, fatty acids, and foreign compounds in liver cells, cytochromes can also be induced by toxic chemicals, pesticides, and cancerogens.

Immunological and PCR-based assays for cytochromes may be used to determine toxicity and turnover rate of experimental medicines. Selective cytotoxic drugs can be designed that interact with a particular cytochrome and trigger cell death, thereby providing potential new treatments for cancer.

WO 02/34922

PCT/US01/42528

Cytochromes can generate free radicals that cause myocardial cell injury and induce endothelial cell damage. In experimental models, alpha-tocopherol and other anti-oxidants suppress generation of free radicals. Glutathione and glutathione peroxidase contribute to natural protection against free radical-induced cell damage. Characterization of all cytochromes will assist development of more efficient anti-oxidants. The sequence provided by the present invention can be used to design specific chemopreventive drugs.

The cytochrome provided herein, as well as other human cytochromes, can be used in a high-throughput drug screen to discover anti-parasitic drugs that inhibit non-human oxygenases but exhibit no toxicity for the human enzymes.

For a further review of the CYP superfamily, see Igarashi *et al.*, *Arch Biochem Biophys* 1997 Mar 1;339(1):85-91; *Med Lett Drugs Ther* 2000 Apr 17;42(1076):35-6 (no authors listed); Fowler *et al.*, *Biochemistry* 2000 Apr 18;39(15):4406-14; Lamb *et al.*, *Chem Biol Interact* 2000 Mar 15;125(3):165-75; Chiba *et al.*, *Xenobiotica* 2000 Feb;30(2):117-29; and Meehan *et al.*, *Am J Hum Genet* 1988 Jan;42(1):26-37.

The CYP superfamily a major target for drug action and development. Accordingly, it is valuable to the field of pharmaceutical development to identify and characterize previously unknown members of the CYP superfamily.

UDP-glucuronosyltransferases

Potential drug interactions involving phase II metabolism are increasingly being recognized. An important group of phase II enzymes involved in drug metabolism are the glucuronosyltransferases, especially the UDP-glucuronosyltransferase ("UGT") superfamily. Members of the UGT superfamily catalyze the enzymatic addition of UDP glucuronic acid as a sugar donor to fat-soluble chemicals, a process which increases their solubility in water and increases their rate of excretion. In mammals, glucuronic acid is the main sugar that is used to prevent the accumulation of waste products of metabolism and fat-soluble chemicals from the environment to toxic levels in the body. Both inducers and inhibitors of glucuronosyltransferases are known and have the potential to affect the plasma concentration and actions of important drugs, including psychotropic drugs.

The UGT superfamily comprises several families of enzymes in several species defined with a nomenclature similar to that used to define members of the CYP superfamily. In animals, yeast, plants and bacteria there are at least 110 distinct known members of the UGT superfamily.

WO 02/34922

PCT/US01/42528

As many as 33 families have been defined, with three families identified in humans. Different UGT families are defined as having <45% amino acid sequence homology; within subfamilies there is approximately 60% homology. The members of the UGT superfamily are part of a further superfamily of UDP glycosyltransferases found in animals, plants and bacteria.

5 The role of phase II enzymes, and of UGT enzymes in particular, is being increasingly recognized as important in psychopharmacology. UGT enzymes conjugate many important psychotropic drugs and are an important source of variability in drug response and drug interactions. For example, the benzodiazepines lorazepam, oxazepam, and temazepam undergo phase II reactions exclusively before being excreted into the urine.

10 Phase II enzymes metabolize and detoxify hazardous substances, such as carcinogens. The expression of genes encoding phase II enzymes is known to be up-regulated by hundreds of agents. For example, oltipraz is known to up-regulate phase II enzyme expression. Studies have demonstrated protection from the cancer-causing effects of carcinogens when selected phase II enzyme inducers are administered prior to the carcinogens. The potential use of phase II enzyme inducers in humans for prevention of cancers related to exposure to carcinogens has prompted
15 studies aimed at understanding their molecular effects. Current biochemical and molecular biological research methodologies can be used to identify and characterize selective phase II enzyme inducers and their targets. Identification of genes responding to cancer chemopreventive agents will facilitate studies of their basic mechanism and provide insights about the relationship
20 between gene regulation, enzyme polymorphism, and carcinogen detoxification.

 Examples of drugs with conjugative metabolism associated with UGT enzymes include amitriptyline, buprenorphine, chlorpromazine, clozapine, codeine, cyproheptadine, dihydrocodeine, doxepin, imipramine, lamotrigine, lorazepam, morphine, nalorphine, naltrexone, temazepam, and valproate.

25 Abnormal activity of phase II enzymes has been implicated in a range of human diseases. For example, Gilbert syndrome is an autosomal dominant disorder caused by mutation in the UGT1 gene, and mutations in the UGT1A1 enzyme have been demonstrated to be responsible for Crigler-Najjar syndrome.

 The UGT superfamily a major target for drug action and development. Accordingly, it is
30 valuable to the field of pharmaceutical development to identify and characterize previously unknown members of the UGT superfamily.

 Drug-metabolizing enzymes, particularly members of the omega-hydroxylase cytochrome P450 drug-metabolizing enzyme subfamily, are a major target for drug action and development.

WO 02/34922

PCT/US01/42528

Accordingly, it is valuable to the field of pharmaceutical development to identify and characterize previously unknown members of this subfamily of drug-metabolizing proteins. The present invention advances the state of the art by providing a previously unidentified human drug-metabolizing proteins that have homology to members of the omega-hydroxylase cytochrome P450 drug-metabolizing enzyme subfamily.

SUMMARY OF THE INVENTION

The present invention is based in part on the identification of amino acid sequences of human drug-metabolizing enzyme peptides and proteins that are related to the omega-hydroxylase cytochrome P450 drug-metabolizing enzyme subfamily, as well as allelic variants and other mammalian orthologs thereof. These unique peptide sequences, and nucleic acid sequences that encode these peptides, can be used as models for the development of human therapeutic targets, aid in the identification of therapeutic proteins, and serve as targets for the development of human therapeutic agents that modulate drug-metabolizing enzyme activity in cells and tissues that express the drug-metabolizing enzyme. Experimental data as provided in Figure 1 indicates expression in humans in the stomach, brain (including infant), endometrial tumors, prostate, kidney, adrenal gland tumors, head/neck, sympathetic trunk, breast, and hepatocellular carcinomas.

DESCRIPTION OF THE FIGURE SHEETS

FIGURE 1 provides the nucleotide sequence of a cDNA molecule that encodes the drug-metabolizing enzyme protein of the present invention. (SEQ ID NO:1) In addition, structure and functional information is provided, such as ATG start, stop and tissue distribution, where available, that allows one to readily determine specific uses of inventions based on this molecular sequence. Experimental data as provided in Figure 1 indicates expression in humans in the stomach, brain (including infant), endometrial tumors, prostate, kidney, adrenal gland tumors, head/neck, sympathetic trunk, breast, and hepatocellular carcinomas.

FIGURE 2 provides the predicted amino acid sequence of the drug-metabolizing enzyme of the present invention. (SEQ ID NO:2) In addition structure and functional information such as protein family, function, and modification sites is provided where available, allowing one to readily determine specific uses of inventions based on this molecular sequence.

WO 02/34922

PCT/US01/42528

FIGURE 3 provides genomic sequences that span the gene encoding the drug-metabolizing enzyme protein of the present invention. (SEQ ID NO:3) In addition structure and functional information, such as intron/exon structure, promoter location, etc., is provided where available, allowing one to readily determine specific uses of inventions based on this molecular sequence. As illustrated in Figure 3, SNPs were identified at 45 different nucleotide positions.

DETAILED DESCRIPTION OF THE INVENTION

General Description

The present invention is based on the sequencing of the human genome. During the sequencing and assembly of the human genome, analysis of the sequence information revealed previously unidentified fragments of the human genome that encode peptides that share structural and/or sequence homology to protein/peptide/domains identified and characterized within the art as being a drug-metabolizing enzyme protein or part of a drug-metabolizing enzyme protein and are related to the omega-hydroxylase cytochrome P450 drug-metabolizing enzyme subfamily. Utilizing these sequences, additional genomic sequences were assembled and transcript and/or cDNA sequences were isolated and characterized. Based on this analysis, the present invention provides amino acid sequences of human drug-metabolizing enzyme peptides and proteins that are related to the omega-hydroxylase cytochrome P450 drug-metabolizing enzyme subfamily, nucleic acid sequences in the form of transcript sequences, cDNA sequences and/or genomic sequences that encode these drug-metabolizing enzyme peptides and proteins, nucleic acid variation (allelic information), tissue distribution of expression, and information about the closest art known protein/peptide/domain that has structural or sequence homology to the drug-metabolizing enzyme of the present invention.

In addition to being previously unknown, the peptides that are provided in the present invention are selected based on their ability to be used for the development of commercially important products and services. Specifically, the present peptides are selected based on homology and/or structural relatedness to known drug-metabolizing enzyme proteins of the omega-hydroxylase cytochrome P450 drug-metabolizing enzyme subfamily and the expression pattern observed. Experimental data as provided in Figure 1 indicates expression in humans in the stomach, brain (including infant), endometrial tumors, prostate, kidney, adrenal gland tumors, head/neck, sympathetic trunk, breast, and hepatocellular carcinomas. The art has clearly established the commercial importance of members of this family of proteins and proteins that

WO 02/34922

PCT/US01/42528

have expression patterns similar to that of the present gene. Some of the more specific features of the peptides of the present invention, and the uses thereof, are described herein, particularly in the Background of the Invention and in the annotation provided in the Figures, and/or are known within the art for each of the known omega-hydroxylase cytochrome P450 family or subfamily of drug-metabolizing enzyme proteins.

Specific Embodiments

Peptide Molecules

The present invention provides nucleic acid sequences that encode protein molecules that have been identified as being members of the drug-metabolizing enzyme family of proteins and are related to the omega-hydroxylase cytochrome P450 drug-metabolizing enzyme subfamily (protein sequences are provided in Figure 2, transcript/cDNA sequences are provided in Figure 1 and genomic sequences are provided in Figure 3). The peptide sequences provided in Figure 2, as well as the obvious variants described herein, particularly allelic variants as identified herein and using the information in Figure 3, will be referred herein as the drug-metabolizing enzyme peptides of the present invention, drug-metabolizing enzyme peptides, or peptides/proteins of the present invention.

The present invention provides isolated peptide and protein molecules that consist of, consist essentially of, or comprise the amino acid sequences of the drug-metabolizing enzyme peptides disclosed in the Figure 2, (encoded by the nucleic acid molecule shown in Figure 1, transcript/cDNA or Figure 3, genomic sequence), as well as all obvious variants of these peptides that are within the art to make and use. Some of these variants are described in detail below.

As used herein, a peptide is said to be "isolated" or "purified" when it is substantially free of cellular material or free of chemical precursors or other chemicals. The peptides of the present invention can be purified to homogeneity or other degrees of purity. The level of purification will be based on the intended use. The critical feature is that the preparation allows for the desired function of the peptide, even if in the presence of considerable amounts of other components (the features of an isolated nucleic acid molecule is discussed below).

In some uses, "substantially free of cellular material" includes preparations of the peptide having less than about 30% (by dry weight) other proteins (i.e., contaminating protein), less than about 20% other proteins, less than about 10% other proteins, or less than about 5% other proteins.

WO 02/34922

PCT/US01/42528

When the peptide is recombinantly produced, it can also be substantially free of culture medium, i.e., culture medium represents less than about 20% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of the peptide in which it is separated from chemical precursors or other chemicals that are involved in its synthesis. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of the drug-metabolizing enzyme peptide having less than about 30% (by dry weight) chemical precursors or other chemicals, less than about 20% chemical precursors or other chemicals, less than about 10% chemical precursors or other chemicals, or less than about 5% chemical precursors or other chemicals.

The isolated drug-metabolizing enzyme peptide can be purified from cells that naturally express it, purified from cells that have been altered to express it (recombinant), or synthesized using known protein synthesis methods. Experimental data as provided in Figure 1 indicates expression in humans in the stomach, brain (including infant), endometrial tumors, prostate, kidney, adrenal gland tumors, head/neck, sympathetic trunk, breast, and hepatocellular carcinomas. For example, a nucleic acid molecule encoding the drug-metabolizing enzyme peptide is cloned into an expression vector, the expression vector introduced into a host cell and the protein expressed in the host cell. The protein can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques. Many of these techniques are described in detail below.

Accordingly, the present invention provides proteins that consist of the amino acid sequences provided in Figure 2 (SEQ ID NO:2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NO:1) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). The amino acid sequence of such a protein is provided in Figure 2. A protein consists of an amino acid sequence when the amino acid sequence is the final amino acid sequence of the protein.

The present invention further provides proteins that consist essentially of the amino acid sequences provided in Figure 2 (SEQ ID NO:2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NO:1) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). A protein consists essentially of an amino acid sequence when such an amino acid sequence is present with only a few additional amino acid residues, for example from about 1 to about 100 or so additional residues, typically from 1 to about 20 additional residues in the final protein.

WO 02/34922

PCT/US01/42528

The present invention further provides proteins that comprise the amino acid sequences provided in Figure 2 (SEQ ID NO:2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NO:1) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). A protein comprises an amino acid sequence when the amino acid sequence is at least part of the final amino acid sequence of the protein. In such a fashion, the protein can be only the peptide or have additional amino acid molecules, such as amino acid residues (contiguous encoded sequence) that are naturally associated with it or heterologous amino acid residues/peptide sequences. Such a protein can have a few additional amino acid residues or can comprise several hundred or more additional amino acids. The preferred classes of proteins that are comprised of the drug-metabolizing enzyme peptides of the present invention are the naturally occurring mature proteins. A brief description of how various types of these proteins can be made/isolated is provided below.

The drug-metabolizing enzyme peptides of the present invention can be attached to heterologous sequences to form chimeric or fusion proteins. Such chimeric and fusion proteins comprise a drug-metabolizing enzyme peptide operatively linked to a heterologous protein having an amino acid sequence not substantially homologous to the drug-metabolizing enzyme peptide. "Operatively linked" indicates that the drug-metabolizing enzyme peptide and the heterologous protein are fused in-frame. The heterologous protein can be fused to the N-terminus or C-terminus of the drug-metabolizing enzyme peptide.

In some uses, the fusion protein does not affect the activity of the drug-metabolizing enzyme peptide *per se*. For example, the fusion protein can include, but is not limited to, enzymatic fusion proteins, for example beta-galactosidase fusions, yeast two-hybrid GAL fusions, poly-His fusions, MYC-tagged, HI-tagged and Ig fusions. Such fusion proteins, particularly poly-His fusions, can facilitate the purification of recombinant drug-metabolizing enzyme peptide. In certain host cells (e.g., mammalian host cells), expression and/or secretion of a protein can be increased by using a heterologous signal sequence.

A chimeric or fusion protein can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different protein sequences are ligated together in-frame in accordance with conventional techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and re-amplified to generate a chimeric gene sequence (see Ausubel *et al.*, *Current*

WO 02/34922

PCT/US01/42528

Protocols in Molecular Biology, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST protein). A drug-metabolizing enzyme peptide-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the drug-metabolizing enzyme peptide.

5 As mentioned above, the present invention also provides and enables obvious variants of the amino acid sequence of the proteins of the present invention, such as naturally occurring mature forms of the peptide, allelic/sequence variants of the peptides, non-naturally occurring recombinantly derived variants of the peptides, and orthologs and paralogs of the peptides. Such variants can readily be generated using art-known techniques in the fields of recombinant nucleic
10 acid technology and protein biochemistry. It is understood, however, that variants exclude any amino acid sequences disclosed prior to the invention.

Such variants can readily be identified/made using molecular techniques and the sequence information disclosed herein. Further, such variants can readily be distinguished from other peptides based on sequence and/or structural homology to the drug-metabolizing enzyme peptides
15 of the present invention. The degree of homology/identity present will be based primarily on whether the peptide is a functional variant or non-functional variant, the amount of divergence present in the paralog family and the evolutionary distance between the orthologs.

To determine the percent identity of two amino acid sequences or two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be
20 introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, at least 30%, 40%, 50%, 60%, 70%, 80%, or 90% or more of the length of a reference sequence is aligned for comparison purposes. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared.
25 When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two
30 sequences.

The comparison of sequences and determination of percent identity and similarity between two sequences can be accomplished using a mathematical algorithm. (*Computational*

WO 02/34922

PCT/US01/42528

Molecular Biology, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part 1*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and

5 *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package (available at <http://www.gcg.com>), using either a Blossom 62 matrix or a PAM250 matrix, and a gap weight

10 of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (Devereux, J., *et al.*, *Nucleic Acids Res.* 12(1):387 (1984)) (available at <http://www.gcg.com>), using a NWSgapdna.CMP matrix and a gap weight of

15 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Myers and W. Miller (CABIOS, 4:11-17 (1989)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid and protein sequences of the present invention can further be used as a

20 "query sequence" to perform a search against sequence databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, *et al.* (*J. Mol. Biol.* 215:403-10 (1990)). BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to the nucleic acid molecules of the invention.

25 BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to the proteins of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.* (*Nucleic Acids Res.* 25(17):3389-3402 (1997)). When utilizing BLAST and gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can

30 be used.

Full-length pre-processed forms, as well as mature processed forms, of proteins that comprise one of the peptides of the present invention can readily be identified as having complete sequence identity to one of the drug-metabolizing enzyme peptides of the present invention as well

WO 02/34922

PCT/US01/42528

as being encoded by the same genetic locus as the drug-metabolizing enzyme peptide provided herein. The gene encoding the novel drug-metabolizing protein of the present invention is located on a genome component that has been mapped to human chromosome 1 (as indicated in Figure 3), which is supported by multiple lines of evidence, such as STS and BAC map data.

5 Allelic variants of a drug-metabolizing enzyme peptide can readily be identified as being a human protein having a high degree (significant) of sequence homology/identity to at least a portion of the drug-metabolizing enzyme peptide as well as being encoded by the same genetic locus as the drug-metabolizing enzyme peptide provided herein. Genetic locus can readily be determined based on the genomic information provided in Figure 3, such as the genomic sequence mapped to the
10 reference human. The gene encoding the novel drug-metabolizing protein of the present invention is located on a genome component that has been mapped to human chromosome 1 (as indicated in Figure 3), which is supported by multiple lines of evidence, such as STS and BAC map data. As used herein, two proteins (or a region of the proteins) have significant homology when the amino acid sequences are typically at least about 70-80%, 80-90%, and more typically at least about 90-
15 95% or more homologous. A significantly homologous amino acid sequence, according to the present invention, will be encoded by a nucleic acid sequence that will hybridize to a drug-metabolizing enzyme peptide encoding nucleic acid molecule under stringent conditions as more fully described below.

Figure 3 provides SNP information that has been found in the gene encoding the drug-
20 metabolizing proteins of the present invention. SNPs, including insertion/deletion variants ("indels"), were identified at 45 different nucleotide positions. Changes in the amino acid sequence caused by these SNPs can readily be determined using the universal genetic code and the protein sequence provided in Figure 2 as a reference. Positioning of each SNP in exons, introns, or outside the ORF can readily be determined using the DNA positions given for each
25 SNP and the start/stop, exon, and intron coordinates given in the features.

Paralogs of a drug-metabolizing enzyme peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the drug-metabolizing enzyme peptide, as being encoded by a gene from humans, and as having similar activity or function. Two proteins will typically be considered paralogs when the amino acid sequences are
30 typically at least about 60% or greater, and more typically at least about 70% or greater homology through a given region or domain. Such paralogs will be encoded by a nucleic acid sequence that will hybridize to a drug-metabolizing enzyme peptide encoding nucleic acid molecule under moderate to stringent conditions as more fully described below.

WO 02/34922

PCT/US01/42528

Orthologs of a drug-metabolizing enzyme peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the drug-metabolizing enzyme peptide as well as being encoded by a gene from another organism. Preferred orthologs will be isolated from mammals, preferably primates, for the development of human therapeutic targets and agents. Such orthologs will be encoded by a nucleic acid sequence that will hybridize to a drug-metabolizing enzyme peptide encoding nucleic acid molecule under moderate to stringent conditions, as more fully described below, depending on the degree of relatedness of the two organisms yielding the proteins.

Non-naturally occurring variants of the drug-metabolizing enzyme peptides of the present invention can readily be generated using recombinant techniques. Such variants include, but are not limited to deletions, additions and substitutions in the amino acid sequence of the drug-metabolizing enzyme peptide. For example, one class of substitutions are conserved amino acid substitution. Such substitutions are those that substitute a given amino acid in a drug-metabolizing enzyme peptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu, and Ile; interchange of the hydroxyl residues Ser and Thr; exchange of the acidic residues Asp and Glu; substitution between the amide residues Asn and Gln; exchange of the basic residues Lys and Arg; and replacements among the aromatic residues Phe and Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in Bowie *et al.*, *Science* 247:1306-1310 (1990).

Variant drug-metabolizing enzyme peptides can be fully functional or can lack function in one or more activities, e.g. ability to bind substrate, ability to phosphorylate substrate, ability to mediate signaling, etc. Fully functional variants typically contain only conservative variation or variation in non-critical residues or in non-critical regions. Figure 2 provides the result of protein analysis and can be used to identify critical domains/regions. Functional variants can also contain substitution of similar amino acids that result in no change or an insignificant change in function. Alternatively, such substitutions may positively or negatively affect function to some degree.

Non-functional variants typically contain one or more non-conservative amino acid substitutions, deletions, insertions, inversions, or truncation or a substitution, insertion, inversion, or deletion in a critical residue or critical region.

Amino acids that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham *et al.*, *Science* 244:1081-1085 (1989)), particularly using the results provided in Figure 2. The latter procedure

WO 02/34922

PCT/US01/42528

introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity such as drug-metabolizing enzyme activity or in assays such as an *in vitro* proliferative activity. Sites that are critical for binding partner/substrate binding can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith *et al.*, *J. Mol. Biol.* 224:899-904 (1992); de Vos *et al. Science* 255:306-312 (1992)).

The present invention further provides fragments of the drug-metabolizing enzyme peptides, in addition to proteins and peptides that comprise and consist of such fragments, particularly those comprising the residues identified in Figure 2. The fragments to which the invention pertains, however, are not to be construed as encompassing fragments that may be disclosed publicly prior to the present invention.

As used herein, a fragment comprises at least 8, 10, 12, 14, 16, or more contiguous amino acid residues from a drug-metabolizing enzyme peptide. Such fragments can be chosen based on the ability to retain one or more of the biological activities of the drug-metabolizing enzyme peptide or could be chosen for the ability to perform a function, e.g. bind a substrate or act as an immunogen. Particularly important fragments are biologically active fragments, peptides that are, for example, about 8 or more amino acids in length. Such fragments will typically comprise a domain or motif of the drug-metabolizing enzyme peptide, e.g., active site, a transmembrane domain or a substrate-binding domain. Further, possible fragments include, but are not limited to, domain or motif containing fragments, soluble peptide fragments, and fragments containing immunogenic structures. Predicted domains and functional sites are readily identifiable by computer programs well known and readily available to those of skill in the art (e.g., PROSITE analysis). The results of one such analysis are provided in Figure 2.

Polypeptides often contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally occurring amino acids. Further, many amino acids, including the terminal amino acids, may be modified by natural processes, such as processing and other post-translational modifications, or by chemical modification techniques well known in the art. Common modifications that occur naturally in drug-metabolizing enzyme peptides are described in basic texts, detailed monographs, and the research literature, and they are well known to those of skill in the art (some of these features are identified in Figure 2).

Known modifications include, but are not limited to, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid

WO 02/34922

PCT/US01/42528

derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent crosslinks, formation of cystine, formation of pyroglutamate, formylation, gamma carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing,

5 phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

Such modifications are well known to those of skill in the art and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and
10 ADP-ribosylation, for instance, are described in most basic texts, such as *Proteins - Structure and Molecular Properties*, 2nd Ed., T.E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this subject, such as by Wold, F., *Posttranslational Covalent Modification of Proteins*, B.C. Johnson, Ed., Academic Press, New York 1-12 (1983); Scifter *et al.* (*Metb. Enzymol.* 182: 626-646 (1990)) and Ratten *et al.* (*Ann. N.Y. Acad. Sci.* 663:48-62 (1992)).

15 Accordingly, the drug-metabolizing enzyme peptides of the present invention also encompass derivatives or analogs in which a substituted amino acid residue is not one encoded by the genetic code, in which a substituent group is included, in which the mature drug-metabolizing enzyme peptide is fused with another compound, such as a compound to increase the half-life of the drug-metabolizing enzyme peptide (for example, polyethylene glycol), or in which the additional
20 amino acids are fused to the mature drug-metabolizing enzyme peptide, such as a leader or secretory sequence or a sequence for purification of the mature drug-metabolizing enzyme peptide or a pro-protein sequence.

Protein/Peptide Uses

25 The proteins of the present invention can be used in substantial and specific assays related to the functional information provided in the Figures; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its binding partner or ligand) in biological fluids; and as markers for tissues in which the corresponding protein is preferentially expressed
30 (either constitutively or at a particular stage of tissue differentiation or development or in a disease state). Where the protein binds or potentially binds to another protein or ligand (such as, for example, in a drug-metabolizing enzyme-effector protein interaction or drug-metabolizing

WO 02/34922

PCT/US01/42528

enzyme-ligand interaction), the protein can be used to identify the binding partner/ligand so as to develop a system to identify inhibitors of the binding interaction. Any or all of these uses are capable of being developed into reagent grade or kit format for commercialization as commercial products.

5 Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology: Guide to Molecular Cloning Techniques", Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987.

10 The potential uses of the peptides of the present invention are based primarily on the source of the protein as well as the class/action of the protein. For example, drug-metabolizing enzymes isolated from humans and their human/mammalian orthologs serve as targets for identifying agents for use in mammalian therapeutic applications, e.g. a human drug, particularly in modulating a biological or pathological response in a cell or tissue that expresses the drug-
15 metabolizing enzyme. Experimental data as provided in Figure 1 indicates that drug-metabolizing enzyme proteins of the present invention are expressed in humans in the stomach, brain (including infant), endometrial tumors, prostate, kidney, adrenal gland tumors, head/neck, sympathetic trunk, breast, and hepatocellular carcinomas, as indicated by virtual northern blot analysis. PCR-based tissue screening panels also indicate expression in the brain. A large
20 percentage of pharmaceutical agents are being developed that modulate the activity of drug-metabolizing enzyme proteins, particularly members of the omega-hydroxylase cytochrome P450 subfamily (see Background of the Invention). The structural and functional information provided in the Background and Figures provide specific and substantial uses for the molecules of the present invention, particularly in combination with the expression information provided in
25 Figure 1. Experimental data as provided in Figure 1 indicates expression in humans in the stomach, brain (including infant), endometrial tumors, prostate, kidney, adrenal gland tumors, head/neck, sympathetic trunk, breast, and hepatocellular carcinomas. Such uses can readily be determined using the information provided herein, that which is known in the art, and routine experimentation.

30 The drug-metabolizing enzyme polypeptides (including variants and fragments that may have been disclosed prior to the present invention) are useful for biological assays related to drug-metabolizing enzymes that are related to members of the omega-hydroxylase cytochrome P450 subfamily. Such assays involve any of the known drug-metabolizing enzyme functions or activities

WO 02/34922

PCT/US01/42528

or properties useful for diagnosis and treatment of drug-metabolizing enzyme-related conditions that are specific for the subfamily of drug-metabolizing enzymes that the one of the present invention belongs to, particularly in cells and tissues that express the drug-metabolizing enzyme. Experimental data as provided in Figure 1 indicates that drug-metabolizing enzyme proteins of the present invention are expressed in humans in the stomach, brain (including infant), endometrial tumors, prostate, kidney, adrenal gland tumors, head/neck, sympathetic trunk, breast, and hepatocellular carcinomas, as indicated by virtual northern blot analysis. PCR-based tissue screening panels also indicate expression in the brain.

The drug-metabolizing enzyme polypeptides are also useful in drug screening assays, in cell-based or cell-free systems. Cell-based systems can be native, i.e., cells that normally express the drug-metabolizing enzyme, as a biopsy or expanded in cell culture. Experimental data as provided in Figure 1 indicates expression in humans in the stomach, brain (including infant), endometrial tumors, prostate, kidney, adrenal gland tumors, head/neck, sympathetic trunk, breast, and hepatocellular carcinomas. In an alternate embodiment, cell-based assays involve recombinant host cells expressing the drug-metabolizing enzyme protein.

The polypeptides can be used to identify compounds that modulate drug-metabolizing enzyme activity of the protein in its natural state or an altered form that causes a specific disease or pathology associated with the drug-metabolizing enzyme. Both the drug-metabolizing enzymes of the present invention and appropriate variants and fragments can be used in high-throughput screens to assay candidate compounds for the ability to bind to the drug-metabolizing enzyme. These compounds can be further screened against a functional drug-metabolizing enzyme to determine the effect of the compound on the drug-metabolizing enzyme activity. Further, these compounds can be tested in animal or invertebrate systems to determine activity/effectiveness. Compounds can be identified that activate (agonist) or inactivate (antagonist) the drug-metabolizing enzyme to a desired degree.

Further, the drug-metabolizing enzyme polypeptides can be used to screen a compound for the ability to stimulate or inhibit interaction between the drug-metabolizing enzyme protein and a molecule that normally interacts with the drug-metabolizing enzyme protein. Such assays typically include the steps of combining the drug-metabolizing enzyme protein with a candidate compound under conditions that allow the drug-metabolizing enzyme protein, or fragment, to interact with the target molecule, and to detect the formation of a complex between the protein and the target or to detect the biochemical consequence of the interaction with the drug-metabolizing enzyme protein and the target.

WO 02/34922

PCT/US01/42528

Candidate compounds include, for example, 1) peptides such as soluble peptides, including Ig-tailed fusion peptides and members of random peptide libraries (see, e.g., Laro *et al.*, *Nature* 354:82-84 (1991); Houghten *et al.*, *Nature* 354:84-86 (1991)) and combinatorial chemistry-derived molecular libraries made of D- and/or L- configuration amino acids; 2) phosphopeptides (e.g., members of random and partially degenerate, directed phosphopeptide libraries, see, e.g., Songyang *et al.*, *Cell* 72:767-778 (1993)); 3) antibodies (e.g., polyclonal, monoclonal, humanized, anti-idiotypic, chimeric, and single chain antibodies as well as Fab, F(ab')₂, Fab expression library fragments, and epitope-binding fragments of antibodies); and 4) small organic and inorganic molecules (e.g., molecules obtained from combinatorial and natural product libraries).

One candidate compound is a soluble fragment of the receptor that competes for substrate binding. Other candidate compounds include mutant drug-metabolizing enzymes or appropriate fragments containing mutations that affect drug-metabolizing enzyme function and thus compete for substrate. Accordingly, a fragment that competes for substrate, for example with a higher affinity, or a fragment that binds substrate but does not allow release, is encompassed by the invention.

Any of the biological or biochemical functions mediated by the drug-metabolizing enzyme can be used as an endpoint assay. These include all of the biochemical or biochemical/biological events described herein, in the references cited herein, incorporated by reference for these endpoint assay targets, and other functions known to those of ordinary skill in the art or that can be readily identified using the information provided in the Figures, particularly Figure 2. Specifically, a biological function of a cell or tissues that expresses the drug-metabolizing enzyme can be assayed. Experimental data as provided in Figure 1 indicates that drug-metabolizing enzyme proteins of the present invention are expressed in humans in the stomach, brain (including infant), endometrial tumors, prostate, kidney, adrenal gland tumors, head/neck, sympathetic trunk, breast, and hepatocellular carcinomas, as indicated by virtual northern blot analysis. PCR-based tissue screening panels also indicate expression in the brain.

Binding and/or activating compounds can also be screened by using chimeric drug-metabolizing enzyme proteins in which the amino terminal extracellular domain, or parts thereof, the entire transmembrane domain or subregions, such as any of the seven transmembrane segments or any of the intracellular or extracellular loops and the carboxy terminal intracellular domain, or parts thereof, can be replaced by heterologous domains or subregions. For example, a substrate-binding region can be used that interacts with a different substrate than that which is recognized by the native drug-metabolizing enzyme. Accordingly, a different set of signal transduction

WO 02/34922

PCT/US01/42528

components is available as an end-point assay for activation. This allows for assays to be performed in other than the specific host cell from which the drug-metabolizing enzyme is derived.

The drug-metabolizing enzyme polypeptides are also useful in competition binding assays in methods designed to discover compounds that interact with the drug-metabolizing enzyme (e.g. binding partners and/or ligands). Thus, a compound is exposed to a drug-metabolizing enzyme polypeptide under conditions that allow the compound to bind or to otherwise interact with the polypeptide. Soluble drug-metabolizing enzyme polypeptide is also added to the mixture. If the test compound interacts with the soluble drug-metabolizing enzyme polypeptide, it decreases the amount of complex formed or activity from the drug-metabolizing enzyme target. This type of assay is particularly useful in cases in which compounds are sought that interact with specific regions of the drug-metabolizing enzyme. Thus, the soluble polypeptide that competes with the target drug-metabolizing enzyme region is designed to contain peptide sequences corresponding to the region of interest.

To perform cell free drug screening assays, it is sometimes desirable to immobilize either the drug-metabolizing enzyme protein, or fragment, or its target molecule to facilitate separation of complexes from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay.

Techniques for immobilizing proteins on matrices can be used in the drug screening assays. In one embodiment, a fusion protein can be provided which adds a domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the cell lysates (e.g., ³⁵S-labeled) and the candidate compound, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads are washed to remove any unbound label, and the matrix immobilized and radiolabel determined directly, or in the supernatant after the complexes are dissociated. Alternatively, the complexes can be dissociated from the matrix, separated by SDS-PAGE, and the level of drug-metabolizing enzyme-binding protein found in the bead fraction quantitated from the gel using standard electrophoretic techniques. For example, either the polypeptide or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin using techniques well known in the art. Alternatively, antibodies reactive with the protein but which do not interfere with binding of the protein to its target molecule can be derivatized to the wells of the plate, and the protein trapped in the wells by antibody conjugation. Preparations of a drug-metabolizing enzyme-binding protein and a candidate

WO 02/34922

PCT/US01/42528

compound are incubated in the drug-metabolizing enzyme protein-presenting wells and the amount of complex trapped in the well can be quantitated. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the drug-metabolizing enzyme protein target molecule, or
 5 which are reactive with drug-metabolizing enzyme protein and compete with the target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the target molecule.

Agents that modulate one of the drug-metabolizing enzymes of the present invention can be identified using one or more of the above assays, alone or in combination. It is generally preferable
 10 to use a cell-based or cell free system first and then confirm activity in an animal or other model system. Such model systems are well known in the art and can readily be employed in this context.

Modulators of drug-metabolizing enzyme protein activity identified according to these drug screening assays can be used to treat a subject with a disorder mediated by the drug-metabolizing enzyme pathway, by treating cells or tissues that express the drug-metabolizing enzyme.
 15 Experimental data as provided in Figure 1 indicates expression in humans in the stomach, brain (including infant), endometrial tumors, prostate, kidney, adrenal gland tumors, head/neck, sympathetic trunk, breast, and hepatocellular carcinomas. These methods of treatment include the steps of administering a modulator of drug-metabolizing enzyme activity in a pharmaceutical composition to a subject in need of such treatment, the modulator being identified as described
 20 herein.

In yet another aspect of the invention, the drug-metabolizing enzyme proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos *et al.* (1993) *Cell* 72:223-232; Madurn *et al.* (1993) *J. Biol. Chem.* 268:12046-12054; Bartel *et al.* (1993) *Biotechniques* 14:920-924; Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins, which bind to or interact with the
 25 drug-metabolizing enzyme and are involved in drug-metabolizing enzyme activity. Such drug-metabolizing enzyme-binding proteins are likely to be drug-metabolizing enzyme inhibitors.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two
 30 different DNA constructs. In one construct, the gene that codes for a drug-metabolizing enzyme protein is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the

WO 02/34922

PCT/US01/42528

activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a drug-metabolizing enzyme-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a
5 transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the drug-metabolizing enzyme protein.

This invention further pertains to novel agents identified by the above-described
10 screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein in an appropriate animal model. For example, an agent identified as described herein (e.g., a drug-metabolizing enzyme-modulating agent, an antisense drug-metabolizing enzyme nucleic acid molecule, a drug-metabolizing enzyme-specific antibody, or a drug-metabolizing enzyme-binding partner) can be used in an animal or other model to
15 determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal or other model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

The drug-metabolizing enzyme proteins of the present invention are also useful to provide a
20 target for diagnosing a disease or predisposition to disease mediated by the peptide. Accordingly, the invention provides methods for detecting the presence, or levels of, the protein (or encoding mRNA) in a cell, tissue, or organism. Experimental data as provided in Figure 1 indicates expression in humans in the stomach, brain (including infant), endometrial tumors, prostate, kidney, adrenal gland tumors, head/neck, sympathetic trunk, breast, and hepatocellular carcinomas. The
25 method involves contacting a biological sample with a compound capable of interacting with the drug-metabolizing enzyme protein such that the interaction can be detected. Such an assay can be provided in a single detection format or a multi-detection format such as an antibody chip array.

One agent for detecting a protein in a sample is an antibody capable of selectively binding to protein. A biological sample includes tissues, cells and biological fluids isolated from a subject, as
30 well as tissues, cells and fluids present within a subject.

The peptides of the present invention also provide targets for diagnosing active protein activity, disease, or predisposition to disease, in a patient having a variant peptide, particularly activities and conditions that are known for other members of the family of proteins to which the

WO 02/34922

PCT/US01/42528

present one belongs. Thus, the peptide can be isolated from a biological sample and assayed for the presence of a genetic mutation that results in aberrant peptide. This includes amino acid substitution, deletion, insertion, rearrangement, (as the result of aberrant splicing events), and inappropriate post-translational modification. Analytic methods include altered electrophoretic mobility, altered tryptic peptide digest, altered drug-metabolizing enzyme activity in cell-based or cell-free assay, alteration in substrate or antibody-binding pattern, altered isoelectric point, direct amino acid sequencing, and any other of the known assay techniques useful for detecting mutations in a protein. Such an assay can be provided in a single detection format or a multi-detection format such as an antibody chip array.

In vitro techniques for detection of peptide include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence using a detection reagent, such as an antibody or protein binding agent. Alternatively, the peptide can be detected *in vivo* in a subject by introducing into the subject a labeled anti-peptide antibody or other types of detection agent. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques. Particularly useful are methods that detect the allelic variant of a peptide expressed in a subject and methods which detect fragments of a peptide in a sample.

The peptides are also useful in pharmacogenomic analysis. Pharmacogenomics deal with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, e.g., Eichelbaum, M. (*Clin. Exp. Pharmacol. Physiol.* 23(10-11):983-985 (1996)), and Linder, M.W. (*Clin. Chem.* 43(2):254-266 (1997)). The clinical outcomes of these variations result in severe toxicity of therapeutic drugs in certain individuals or therapeutic failure of drugs in certain individuals as a result of individual variation in metabolism. Thus, the genotype of the individual can determine the way a therapeutic compound acts on the body or the way the body metabolizes the compound. Further, the activity of drug metabolizing enzymes effects both the intensity and duration of drug action. Thus, the pharmacogenomics of the individual permit the selection of effective compounds and effective dosages of such compounds for prophylactic or therapeutic treatment based on the individual's genotype. The discovery of genetic polymorphisms in some drug metabolizing enzymes has explained why some patients do not obtain the expected drug effects, show an exaggerated drug effect, or experience serious toxicity from standard drug dosages. Polymorphisms can be expressed in the phenotype of the extensive metabolizer and the phenotype of the poor metabolizer. Accordingly, genetic polymorphism may lead to allelic protein variants of the drug-metabolizing enzyme protein in which one or more of the

WO 02/34922

PCT/US01/42528

drug-metabolizing enzyme functions in one population is different from those in another population. The peptides thus allow a target to ascertain a genetic predisposition that can affect treatment modality. Thus, in a ligand-based treatment, polymorphism may give rise to amino terminal extracellular domains and/or other substrate-binding regions that are more or less active in substrate binding, and drug-metabolizing enzyme activation. Accordingly, substrate dosage would necessarily be modified to maximize the therapeutic effect within a given population containing a polymorphism. As an alternative to genotyping, specific polymorphic peptides could be identified.

The peptides are also useful for treating a disorder characterized by an absence of, inappropriate, or unwanted expression of the protein. Experimental data as provided in Figure 1 indicates expression in humans in the stomach, brain (including infant), endometrial tumors, prostate, kidney, adrenal gland tumors, head/neck, sympathetic trunk, breast, and hepatocellular carcinomas. Accordingly, methods for treatment include the use of the drug-metabolizing enzyme protein or fragments.

15 Antibodies

The invention also provides antibodies that selectively bind to one of the peptides of the present invention, a protein comprising such a peptide, as well as variants and fragments thereof. As used herein, an antibody selectively binds a target peptide when it binds the target peptide and does not significantly bind to unrelated proteins. An antibody is still considered to selectively bind a peptide even if it also binds to other proteins that are not substantially homologous with the target peptide so long as such proteins share homology with a fragment or domain of the peptide target of the antibody. In this case, it would be understood that antibody binding to the peptide is still selective despite some degree of cross-reactivity.

As used herein, an antibody is defined in terms consistent with that recognized within the art: they are multi-subunit proteins produced by a mammalian organism in response to an antigen challenge. The antibodies of the present invention include polyclonal antibodies and monoclonal antibodies, as well as fragments of such antibodies, including, but not limited to, Fab or F(ab')₂ and Fv fragments.

Many methods are known for generating and/or identifying antibodies to a given target peptide. Several such methods are described by Harlow, Antibodies, Cold Spring Harbor Press, (1989).

WO 02/34922

PCT/US01/42528

In general, to generate antibodies, an isolated peptide is used as an immunogen and is administered to a mammalian organism, such as a rat, rabbit or mouse. The full-length protein, an antigenic peptide fragment or a fusion protein can be used. Particularly important fragments are those covering functional domains, such as the domains identified in Figure 2, and domain of sequence homology or divergence amongst the family, such as those that can readily be identified using protein alignment methods and as presented in the Figures.

Antibodies are preferably prepared from regions or discrete fragments of the drug-metabolizing enzyme proteins. Antibodies can be prepared from any region of the peptide as described herein. However, preferred regions will include those involved in function/activity and/or drug-metabolizing enzyme/binding partner interaction. Figure 2 can be used to identify particularly important regions while sequence alignment can be used to identify conserved and unique sequence fragments.

An antigenic fragment will typically comprise at least 8 contiguous amino acid residues. The antigenic peptide can comprise, however, at least 10, 12, 14, 16 or more amino acid residues. Such fragments can be selected on a physical property, such as fragments correspond to regions that are located on the surface of the protein, e.g., hydrophilic regions or can be selected based on sequence uniqueness (see Figure 2).

Detection on an antibody of the present invention can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luciferase; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I , ^{131}I , ^{35}S or ^3H .

Antibody Uses

The antibodies can be used to isolate one of the proteins of the present invention by standard techniques, such as affinity chromatography or immunoprecipitation. The antibodies can facilitate the purification of the natural protein from cells and recombinantly produced protein expressed in

WO 02/34922

PCT/US01/42528

host cells. In addition, such antibodies are useful to detect the presence of one of the proteins of the present invention in cells or tissues to determine the pattern of expression of the protein among various tissues in an organism and over the course of normal development. Experimental data as provided in Figure 1 indicates that drug-metabolizing enzyme proteins of the present invention are
5 expressed in humans in the stomach, brain (including infant), endometrial tumors, prostate, kidney, adrenal gland tumors, head/neck, sympathetic trunk, breast, and hepatocellular carcinomas, as indicated by virtual northern blot analysis. PCR-based tissue screening panels also indicate expression in the brain. Further, such antibodies can be used to detect protein *in situ*, *in vitro*, or in a cell lysate or supernatant in order to evaluate the abundance and pattern of expression. Also, such
10 antibodies can be used to assess abnormal tissue distribution or abnormal expression during development or progression of a biological condition. Antibody detection of circulating fragments of the full length protein can be used to identify turnover.

Further, the antibodies can be used to assess expression in disease states such as in active stages of the disease or in an individual with a predisposition toward disease related to the protein's
15 function. When a disorder is caused by an inappropriate tissue distribution, developmental expression, level of expression of the protein, or expressed/processed form, the antibody can be prepared against the normal protein. Experimental data as provided in Figure 1 indicates expression in humans in the stomach, brain (including infant), endometrial tumors, prostate, kidney, adrenal gland tumors, head/neck, sympathetic trunk, breast, and hepatocellular carcinomas. If a disorder is
20 characterized by a specific mutation in the protein, antibodies specific for this mutant protein can be used to assay for the presence of the specific mutant protein.

The antibodies can also be used to assess normal and aberrant subcellular localization of cells in the various tissues in an organism. Experimental data as provided in Figure 1 indicates
25 expression in humans in the stomach, brain (including infant), endometrial tumors, prostate, kidney, adrenal gland tumors, head/neck, sympathetic trunk, breast, and hepatocellular carcinomas. The diagnostic uses can be applied, not only in genetic testing, but also in monitoring a treatment modality. Accordingly, where treatment is ultimately aimed at correcting expression level or the presence of aberrant sequence and aberrant tissue distribution or developmental expression, antibodies directed against the protein or relevant fragments can be used to monitor therapeutic
30 efficacy.

Additionally, antibodies are useful in pharmacogenomic analysis. Thus, antibodies prepared against polymorphic proteins can be used to identify individuals that require modified treatment modalities. The antibodies are also useful as diagnostic tools as an immunological marker for

WO 02/34922

PCT/US01/42528

aberrant protein analyzed by electrophoretic mobility, isoelectric point, tryptic peptide digest, and other physical assays known to those in the art.

The antibodies are also useful for tissue typing. Experimental data as provided in Figure 1 indicates expression in humans in the stomach, brain (including infant), endometrial tumors, prostate, kidney, adrenal gland tumors, head/neck, sympathetic trunk, breast, and hepatocellular carcinomas. Thus, where a specific protein has been correlated with expression in a specific tissue, antibodies that are specific for this protein can be used to identify a tissue type.

The antibodies are also useful for inhibiting protein function, for example, blocking the binding of the drug-metabolizing enzyme peptide to a binding partner such as a substrate. These uses can also be applied in a therapeutic context in which treatment involves inhibiting the protein's function. An antibody can be used, for example, to block binding, thus modulating (agonizing or antagonizing) the peptide's activity. Antibodies can be prepared against specific fragments containing sites required for function or against intact protein that is associated with a cell or cell membrane. See Figure 2 for structural information relating to the proteins of the present invention.

The invention also encompasses kits for using antibodies to detect the presence of a protein in a biological sample. The kit can comprise antibodies such as a labeled or labelable antibody and a compound or agent for detecting protein in a biological sample; means for determining the amount of protein in the sample; means for comparing the amount of protein in the sample with a standard; and instructions for use. Such a kit can be supplied to detect a single protein or epitope or can be configured to detect one of a multitude of epitopes, such as in an antibody detection array. Arrays are described in detail below for nucleic acid arrays and similar methods have been developed for antibody arrays.

Nucleic Acid Molecules

The present invention further provides isolated nucleic acid molecules that encode a drug-metabolizing enzyme peptide or protein of the present invention (cDNA, transcript and genomic sequence). Such nucleic acid molecules will consist of, consist essentially of, or comprise a nucleotide sequence that encodes one of the drug-metabolizing enzyme peptides of the present invention, an allelic variant thereof, or an ortholog or paralog thereof.

As used herein, an "isolated" nucleic acid molecule is one that is separated from other nucleic acid present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences that naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends

WO 02/34922

PCT/US01/42528

of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. However, there can be some flanking nucleotide sequences, for example up to about 5KB, 4KB, 3KB, 2KB, or 1KB or less, particularly contiguous peptide encoding sequences and peptide encoding sequences within the same gene but separated by introns in the genomic sequence. The
 5 important point is that the nucleic acid is isolated from remote and unimportant flanking sequences such that it can be subjected to the specific manipulations described herein such as recombinant expression, preparation of probes and primers, and other uses specific to the nucleic acid sequences.

Moreover, an "isolated" nucleic acid molecule, such as a transcript/cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant
 10 techniques, or chemical precursors or other chemicals when chemically synthesized. However, the nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered isolated.

For example, recombinant DNA molecules contained in a vector are considered isolated. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in
 15 heterologous host cells or purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of the isolated DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

Accordingly, the present invention provides nucleic acid molecules that consist of the
 20 nucleotide sequence shown in Figure 1 or 3 (SEQ ID NO:1, transcript sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ ID NO:2. A nucleic acid molecule consists of a nucleotide sequence when the nucleotide sequence is the complete nucleotide sequence of the nucleic acid molecule.

The present invention further provides nucleic acid molecules that consist essentially of the
 25 nucleotide sequence shown in Figure 1 or 3 (SEQ ID NO:1, transcript sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ ID NO:2. A nucleic acid molecule consists essentially of a nucleotide sequence when such a nucleotide sequence is present with only a few additional nucleic acid residues in the final nucleic acid molecule.

The present invention further provides nucleic acid molecules that comprise the nucleotide
 30 sequences shown in Figure 1 or 3 (SEQ ID NO:1, transcript sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ ID NO:2. A nucleic acid molecule comprises a nucleotide sequence when the nucleotide sequence is at

WO 02/34922

PCT/US01/42528

least part of the final nucleotide sequence of the nucleic acid molecule. In such a fashion, the nucleic acid molecule can be only the nucleotide sequence or have additional nucleic acid residues, such as nucleic acid residues that are naturally associated with it or heterologous nucleotide sequences. Such a nucleic acid molecule can have a few additional nucleotides or can comprises
5 several hundred or more additional nucleotides. A brief description of how various types of these nucleic acid molecules can be readily made/isolated is provided below.

In Figures 1 and 3, both coding and non-coding sequences are provided. Because of the source of the present invention, humans genomic sequence (Figure 3) and cDNA/transcript sequences (Figure 1), the nucleic acid molecules in the Figures will contain genomic intronic
10 sequences, 5' and 3' non-coding sequences, gene regulatory regions and non-coding intergenic sequences. In general such sequence features are either noted in Figures 1 and 3 or can readily be identified using computational tools known in the art. As discussed below, some of the non-coding regions, particularly gene regulatory elements such as promoters, are useful for a variety of purposes, e.g. control of heterologous gene expression, target for identifying gene activity
15 modulating compounds, and are particularly claimed as fragments of the genomic sequence provided herein.

The isolated nucleic acid molecules can encode the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature peptide (when the mature form has more than one peptide chain, for instance). Such sequences may play a role in processing of a
20 protein from precursor to a mature form, facilitate protein trafficking, prolong or shorten protein half-life or facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in situ*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

As mentioned above, the isolated nucleic acid molecules include, but are not limited to, the
25 sequence encoding the drug-metabolizing enzyme peptide alone, the sequence encoding the mature peptide and additional coding sequences, such as a leader or secretory sequence (e.g., a pre-pro or pro-protein sequence), the sequence encoding the mature peptide, with or without the additional coding sequences, plus additional non-coding sequences, for example introns and non-coding 5' and 3' sequences such as transcribed but non-translated sequences that play a role in transcription,
30 mRNA processing (including splicing and polyadenylation signals), ribosome binding and stability of mRNA. In addition, the nucleic acid molecule may be fused to a marker sequence encoding, for example, a peptide that facilitates purification.

WO 02/34922

PCT/US01/42528

Isolated nucleic acid molecules can be in the form of RNA, such as mRNA, or in the form of DNA, including cDNA and genomic DNA obtained by cloning or produced by chemical synthetic techniques or by a combination thereof. The nucleic acid, especially DNA, can be double-stranded or single-stranded. Single-stranded nucleic acid can be the coding strand (sense strand) or the non-coding strand (anti-sense strand).

The invention further provides nucleic acid molecules that encode fragments of the peptides of the present invention as well as nucleic acid molecules that encode obvious variants of the drug-metabolizing enzyme proteins of the present invention that are described above. Such nucleic acid molecules may be naturally occurring, such as allelic variants (same locus), paralogs (different locus), and orthologs (different organism), or may be constructed by recombinant DNA methods or by chemical synthesis. Such non-naturally occurring variants may be made by mutagenesis techniques, including those applied to nucleic acid molecules, cells, or organisms. Accordingly, as discussed above, the variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions.

The present invention further provides non-coding fragments of the nucleic acid molecules provided in Figures 1 and 3. Preferred non-coding fragments include, but are not limited to, promoter sequences, enhancer sequences, gene modulating sequences and gene termination sequences. Such fragments are useful in controlling heterologous gene expression and in developing screens to identify gene-modulating agents. A promoter can readily be identified as being 5' to the ATG start site in the genomic sequence provided in Figure 3.

A fragment comprises a contiguous nucleotide sequence greater than 12 or more nucleotides. Further, a fragment could be at least 30, 40, 50, 100, 250 or 500 nucleotides in length. The length of the fragment will be based on its intended use. For example, the fragment can encode epitope bearing regions of the peptide, or can be useful as DNA probes and primers. Such fragments can be isolated using the known nucleotide sequence to synthesize an oligonucleotide probe. A labeled probe can then be used to screen a cDNA library, genomic DNA library, or mRNA to isolate nucleic acid corresponding to the coding region. Further, primers can be used in PCR reactions to clone specific regions of gene.

A probe/primer typically comprises substantially a purified oligonucleotide or oligonucleotide pair. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 20, 25, 40, 50 or more consecutive nucleotides.

WO 02/34922

PCT/US01/42528

Orthologs, homologs, and allelic variants can be identified using methods well known in the art. As described in the Peptide Section, these variants comprise a nucleotide sequence encoding a peptide that is typically 60-70%, 70-80%, 80-90%, and more typically at least about 90-95% or more homologous to the nucleotide sequence shown in the Figure sheets or a fragment of this sequence. Such nucleic acid molecules can readily be identified as being able to hybridize under moderate to stringent conditions, to the nucleotide sequence shown in the Figure sheets or a fragment of the sequence. Allelic variants can readily be determined by genetic locus of the encoding gene. The gene encoding the novel drug-metabolizing protein of the present invention is located on a genome component that has been mapped to human chromosome 1 (as indicated in Figure 3), which is supported by multiple lines of evidence, such as STS and BAC map data.

Figure 3 provides SNP information that has been found in the gene encoding the drug-metabolizing proteins of the present invention. SNPs, including insertion/deletion variants ("indels"), were identified at 45 different nucleotide positions. Changes in the amino acid sequence caused by these SNPs can readily be determined using the universal genetic code and the protein sequence provided in Figure 2 as a reference. Positioning of each SNP in exons, introns, or outside the ORF can readily be determined using the DNA positions given for each SNP and the start/stop, exon, and intron coordinates given in the features.

As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences encoding a peptide at least 60-70% homologous to each other typically remain hybridized to each other. The conditions can be such that sequences at least about 60%, at least about 70%, or at least about 80% or more homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. One example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65C. Examples of moderate to low stringency hybridization conditions are well known in the art.

Nucleic Acid Molecule Uses

The nucleic acid molecules of the present invention are useful for probes, primers, chemical intermediates, and in biological assays. The nucleic acid molecules are useful as a hybridization probe for messenger RNA, transcript/cDNA and genomic DNA to isolate full-length cDNA and

WO 02/34922

PCT/US01/42528

genomic clones encoding the peptide described in Figure 2 and to isolate cDNA and genomic clones that correspond to variants (alleles, orthologs, etc.) producing the same or related peptides shown in Figure 2. As illustrated in Figure 3, SNPs were identified at 45 different nucleotide positions.

5 The probe can correspond to any sequence along the entire length of the nucleic acid molecules provided in the Figures. Accordingly, it could be derived from 5' noncoding regions, the coding region, and 3' noncoding regions. However, as discussed, fragments are not to be construed as encompassing fragments disclosed prior to the present invention.

10 The nucleic acid molecules are also useful as primers for PCR to amplify any given region of a nucleic acid molecule and are useful to synthesize antisense molecules of desired length and sequence.

15 The nucleic acid molecules are also useful for constructing recombinant vectors. Such vectors include expression vectors that express a portion of, or all of, the peptide sequences. Vectors also include insertion vectors, used to integrate into another nucleic acid molecule sequence, such as into the cellular genome, to alter *in situ* expression of a gene and/or gene product. For example, an endogenous coding sequence can be replaced via homologous recombination with all or part of the coding region containing one or more specifically introduced mutations.

 The nucleic acid molecules are also useful for expressing antigenic portions of the proteins.

20 The nucleic acid molecules are also useful as probes for determining the chromosomal positions of the nucleic acid molecules by means of *in situ* hybridization methods. The gene encoding the novel drug-metabolizing protein of the present invention is located on a genome component that has been mapped to human chromosome 1 (as indicated in Figure 3), which is supported by multiple lines of evidence, such as STS and BAC map data.

25 The nucleic acid molecules are also useful in making vectors containing the gene regulatory regions of the nucleic acid molecules of the present invention.

 The nucleic acid molecules are also useful for designing ribozymes corresponding to all, or a part, of the mRNA produced from the nucleic acid molecules described herein.

 The nucleic acid molecules are also useful for making vectors that express part, or all, of the peptides.

30 The nucleic acid molecules are also useful for constructing host cells expressing a part, or all, of the nucleic acid molecules and peptides.

 The nucleic acid molecules are also useful for constructing transgenic animals expressing all, or a part, of the nucleic acid molecules and peptides.

WO 02/24922

PCT/US01/42528

The nucleic acid molecules are also useful as hybridization probes for determining the presence, level, form and distribution of nucleic acid expression. Experimental data as provided in Figure 1 indicates that drug-metabolizing enzyme proteins of the present invention are expressed in humans in the stomach, brain (including infant), endometrial tumors, prostate, kidney, adrenal gland tumors, head/neck, sympathetic trunk, breast, and hepatocellular carcinomas, as indicated by virtual northern blot analysis. PCR-based tissue screening panels also indicate expression in the brain. Accordingly, the probes can be used to detect the presence of, or to determine levels of, a specific nucleic acid molecule in cells, tissues, and in organisms. The nucleic acid whose level is determined can be DNA or RNA. Accordingly, probes corresponding to the peptides described herein can be used to assess expression and/or gene copy number in a given cell, tissue, or organism. These uses are relevant for diagnosis of disorders involving an increase or decrease in drug-metabolizing enzyme protein expression relative to normal results.

In vitro techniques for detection of mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detecting DNA include Southern hybridizations and *in situ* hybridization.

Probes can be used as a part of a diagnostic test kit for identifying cells or tissues that express a drug-metabolizing enzyme protein, such as by measuring a level of a drug-metabolizing enzyme-encoding nucleic acid in a sample of cells from a subject e.g., mRNA or genomic DNA, or determining if a drug-metabolizing enzyme gene has been mutated. Experimental data as provided in Figure 1 indicates that drug-metabolizing enzyme proteins of the present invention are expressed in humans in the stomach, brain (including infant), endometrial tumors, prostate, kidney, adrenal gland tumors, head/neck, sympathetic trunk, breast, and hepatocellular carcinomas, as indicated by virtual northern blot analysis. PCR-based tissue screening panels also indicate expression in the brain.

Nucleic acid expression assays are useful for drug screening to identify compounds that modulate drug-metabolizing enzyme nucleic acid expression.

The invention thus provides a method for identifying a compound that can be used to treat a disorder associated with nucleic acid expression of the drug-metabolizing enzyme gene, particularly biological and pathological processes that are mediated by the drug-metabolizing enzyme in cells and tissues that express it. Experimental data as provided in Figure 1 indicates expression in humans in the stomach, brain (including infant), endometrial tumors, prostate, kidney, adrenal gland tumors, head/neck, sympathetic trunk, breast, and hepatocellular carcinomas. The method typically includes assaying the ability of the compound to modulate the expression of the drug-metabolizing enzyme

WO 02/34922

PCT/US01/42528

nucleic acid and thus identifying a compound that can be used to treat a disorder characterized by undesired drug-metabolizing enzyme nucleic acid expression. The assays can be performed in cell-based and cell-free systems. Cell-based assays include cells naturally expressing the drug-metabolizing enzyme nucleic acid or recombinant cells genetically engineered to express specific nucleic acid sequences.

Thus, modulators of drug-metabolizing enzyme gene expression can be identified in a method wherein a cell is contacted with a candidate compound and the expression of mRNA determined. The level of expression of drug-metabolizing enzyme mRNA in the presence of the candidate compound is compared to the level of expression of drug-metabolizing enzyme mRNA in the absence of the candidate compound. The candidate compound can then be identified as a modulator of nucleic acid expression based on this comparison and be used, for example to treat a disorder characterized by aberrant nucleic acid expression. When expression of mRNA is statistically significantly greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of nucleic acid expression. When nucleic acid expression is statistically significantly less in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of nucleic acid expression.

The invention further provides methods of treatment, with the nucleic acid as a target, using a compound identified through drug screening as a gene modulator to modulate drug-metabolizing enzyme nucleic acid expression in cells and tissues that express the drug-metabolizing enzyme. Experimental data as provided in Figure 1 indicates that drug-metabolizing enzyme proteins of the present invention are expressed in humans in the stomach, brain (including infant), endometrial tumors, prostate, kidney, adrenal gland tumors, head/neck, sympathetic trunk, breast, and hepatocellular carcinomas, as indicated by virtual northern blot analysis. PCR-based tissue screening panels also indicate expression in the brain. Modulation includes both up-regulation (i.e. activation or agonization) or down-regulation (suppression or antagonization) or nucleic acid expression.

Alternatively, a modulator for drug-metabolizing enzyme nucleic acid expression can be a small molecule or drug identified using the screening assays described herein as long as the drug or small molecule inhibits the drug-metabolizing enzyme nucleic acid expression in the cells and tissues that express the protein. Experimental data as provided in Figure 1 indicates expression in humans in the stomach, brain (including infant), endometrial tumors, prostate, kidney, adrenal gland tumors, head/neck, sympathetic trunk, breast, and hepatocellular carcinomas.

WO 02/34922

PCT/US01/42528

The nucleic acid molecules are also useful for monitoring the effectiveness of modulating compounds on the expression or activity of the drug-metabolizing enzyme gene in clinical trials or in a treatment regimen. Thus, the gene expression pattern can serve as a barometer for the continuing effectiveness of treatment with the compound, particularly with compounds to which a patient can develop resistance. The gene expression pattern can also serve as a marker indicative of a physiological response of the affected cells to the compound. Accordingly, such monitoring would allow either increased administration of the compound or the administration of alternative compounds to which the patient has not become resistant. Similarly, if the level of nucleic acid expression falls below a desirable level, administration of the compound could be commensurately decreased.

The nucleic acid molecules are also useful in diagnostic assays for qualitative changes in drug-metabolizing enzyme nucleic acid expression, and particularly in qualitative changes that lead to pathology. The nucleic acid molecules can be used to detect mutations in drug-metabolizing enzyme genes and gene expression products such as mRNA. The nucleic acid molecules can be used as hybridization probes to detect naturally occurring genetic mutations in the drug-metabolizing enzyme gene and thereby to determine whether a subject with the mutation is at risk for a disorder caused by the mutation. Mutations include deletion, addition, or substitution of one or more nucleotides in the gene, chromosomal rearrangement, such as inversion or transposition, modification of genomic DNA, such as aberrant methylation patterns or changes in gene copy number, such as amplification. Detection of a mutated form of the drug-metabolizing enzyme gene associated with a dysfunction provides a diagnostic tool for an active disease or susceptibility to disease when the disease results from overexpression, underexpression, or altered expression of a drug-metabolizing enzyme protein.

Individuals carrying mutations in the drug-metabolizing enzyme gene can be detected at the nucleic acid level by a variety of techniques. Figure 3 provides SNP information that has been found in the gene encoding the drug-metabolizing proteins of the present invention. SNPs, including insertion/deletion variants ("indels"), were identified at 45 different nucleotide positions. Changes in the amino acid sequence caused by these SNPs can readily be determined using the universal genetic code and the protein sequence provided in Figure 2 as a reference. Positioning of each SNP in exons, introns, or outside the ORF can readily be determined using the DNA positions given for each SNP and the start/stop, exon, and intron coordinates given in the features. The gene encoding the novel drug-metabolizing protein of the present invention is located on a genome component that has been mapped to human chromosome 1 (as indicated in Figure 3), which is supported by

WO 02/34922

PCT/US01/42528

multiple lines of evidence, such as STS and BAC map data. Genomic DNA can be analyzed directly or can be amplified by using PCR prior to analysis. RNA or cDNA can be used in the same way. In some uses, detection of the mutation involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g. U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran *et al.*, *Science* 241:1077-1080 (1988); and Nakazawa *et al.*, *PNAS* 91:360-364 (1994)), the latter of which can be particularly useful for detecting point mutations in the gene (see Abrahava *et al.*, *Nucleic Acids Res.* 23:675-682 (1995)). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a gene under conditions such that hybridization and amplification of the gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. Deletions and insertions can be detected by a change in size of the amplified product compared to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to normal RNA or antisense DNA sequences.

Alternatively, mutations in a drug-metabolizing enzyme gene can be directly identified, for example, by alterations in restriction enzyme digestion patterns determined by gel electrophoresis.

Further, sequence-specific ribozymes (U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site. Perfectly matched sequences can be distinguished from mismatched sequences by nuclease cleavage digestion assays or by differences in melting temperature.

Sequence changes at specific locations can also be assessed by nuclease protection assays such as RNase and S1 protection or the chemical cleavage method. Furthermore, sequence differences between a mutant drug-metabolizing enzyme gene and a wild-type gene can be determined by direct DNA sequencing. A variety of automated sequencing procedures can be utilized when performing the diagnostic assays (Naeve, C.W., (1995) *Biotechniques* 19:448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen *et al.*, *Adv. Chromatogr.* 36:127-162 (1996); and Griffin *et al.*, *Appl. Biochem. Biotechnol.* 38:147-159 (1993)).

Other methods for detecting mutations in the gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA duplexes (Myers *et al.*, *Science* 230:1242 (1985)); Cotton *et al.*, *PNAS* 85:4397 (1988); Salceba *et al.*, *Meth.*

WO 02/24922

PCT/US01/42528

Enzymol. 217:286-295 (1992)), electrophoretic mobility of mutant and wild type nucleic acid is compared (Orita *et al.*, *PNAS* 86:2766 (1989); Cotton *et al.*, *Mutat. Res.* 285:125-144 (1993); and Hayashi *et al.*, *Genet. Anal. Tech. Appl.* 9:73-79 (1992)), and movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (Myers *et al.*, *Nature* 313:495 (1985)). Examples of other techniques for detecting point mutations include selective oligonucleotide hybridization, selective amplification, and selective primer extension.

The nucleic acid molecules are also useful for testing an individual for a genotype that while not necessarily causing the disease, nevertheless affects the treatment modality. Thus, the nucleic acid molecules can be used to study the relationship between an individual's genotype and the individual's response to a compound used for treatment (pharmacogenomic relationship). Accordingly, the nucleic acid molecules described herein can be used to assess the mutation content of the drug-metabolizing enzyme gene in an individual in order to select an appropriate compound or dosage regimen for treatment. Figure 3 provides SNP information that has been found in the gene encoding the drug-metabolizing proteins of the present invention. SNPs, including insertion/deletion variants ("indels"), were identified at 45 different nucleotide positions. Changes in the amino acid sequence caused by these SNPs can readily be determined using the universal genetic code and the protein sequence provided in Figure 2 as a reference. Positioning of each SNP in exons, introns, or outside the ORF can readily be determined using the DNA positions given for each SNP and the start/stop, exon, and intron coordinates given in the features.

Thus nucleic acid molecules displaying genetic variations that affect treatment provide a diagnostic target that can be used to tailor treatment in an individual. Accordingly, the production of recombinant cells and animals containing these polymorphisms allow effective clinical design of treatment compounds and dosage regimens.

The nucleic acid molecules are thus useful as antisense constructs to control drug-metabolizing enzyme gene expression in cells, tissues, and organisms. A DNA antisense nucleic acid molecule is designed to be complementary to a region of the gene involved in transcription, preventing transcription and hence production of drug-metabolizing enzyme protein. An antisense RNA or DNA nucleic acid molecule would hybridize to the mRNA and thus block translation of mRNA into drug-metabolizing enzyme protein.

Alternatively, a class of antisense molecules can be used to inactivate mRNA in order to decrease expression of drug-metabolizing enzyme nucleic acid. Accordingly, these molecules can treat a disorder characterized by abnormal or undesired drug-metabolizing enzyme nucleic acid

WO 02/34922

PCT/US01/42528

expression. This technique involves cleavage by means of ribozymes containing nucleotide sequences complementary to one or more regions in the mRNA that attenuate the ability of the mRNA to be translated. Possible regions include coding regions and particularly coding regions corresponding to the catalytic and other functional activities of the drug-metabolizing enzyme protein, such as substrate binding.

The nucleic acid molecules also provide vectors for gene therapy in patients containing cells that are aberrant in drug-metabolizing enzyme gene expression. Thus, recombinant cells, which include the patient's cells that have been engineered *ex vivo* and returned to the patient, are introduced into an individual where the cells produce the desired drug-metabolizing enzyme protein to treat the individual.

The invention also encompasses kits for detecting the presence of a drug-metabolizing enzyme nucleic acid in a biological sample. Experimental data as provided in Figure 1 indicates that drug-metabolizing enzyme proteins of the present invention are expressed in humans in the stomach, brain (including infant), endometrial tumors, prostate, kidney, adrenal gland tumors, head/neck, sympathetic trunk, breast, and hepatocellular carcinomas, as indicated by virtual northern blot analysis. PCR-based tissue screening panels also indicate expression in the brain. For example, the kit can comprise reagents such as a labeled or labelable nucleic acid or agent capable of detecting drug-metabolizing enzyme nucleic acid in a biological sample; means for determining the amount of drug-metabolizing enzyme nucleic acid in the sample; and means for comparing the amount of drug-metabolizing enzyme nucleic acid in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect drug-metabolizing enzyme protein mRNA or DNA.

Nucleic Acid Arrays

The present invention further provides nucleic acid detection kits, such as arrays or microarrays of nucleic acid molecules that are based on the sequence information provided in Figures 1 and 3 (SEQ ID NOS:1 and 3).

As used herein "Arrays" or "Microarrays" refers to an array of distinct polynucleotides or oligonucleotides synthesized on a substrate, such as paper, nylon or other type of membrane, filter, chip, glass slide, or any other suitable solid support. In one embodiment, the microarray is prepared and used according to the methods described in US Patent 5,837,832, Chee *et al.*, PCT application W095/11995 (Chee *et al.*), Lockhart, D. J. *et al.* (1996; Nat. Biotech. 14: 1675-1680)

WO 02/34922

PCT/US01/42528

and Schena, M. *et al.* (1996; Proc. Natl. Acad. Sci. 93: 10614-10619), all of which are incorporated herein in their entirety by reference. In other embodiments, such arrays are produced by the methods described by Brown *et al.*, US Patent No. 5,807,522.

The microarray or detection kit is preferably composed of a large number of unique, single-stranded nucleic acid sequences, usually either synthetic antisense oligonucleotides or fragments of cDNAs, fixed to a solid support. The oligonucleotides are preferably about 6-60 nucleotides in length, more preferably 15-30 nucleotides in length, and most preferably about 20-25 nucleotides in length. For a certain type of microarray or detection kit, it may be preferable to use oligonucleotides that are only 7-20 nucleotides in length. The microarray or detection kit may contain oligonucleotides that cover the known 5', or 3', sequence, sequential oligonucleotides that cover the full length sequence; or unique oligonucleotides selected from particular areas along the length of the sequence. Polynucleotides used in the microarray or detection kit may be oligonucleotides that are specific to a gene or genes of interest.

In order to produce oligonucleotides to a known sequence for a microarray or detection kit, the gene(s) of interest (or an ORF identified from the contigs of the present invention) is typically examined using a computer algorithm which starts at the 5' or at the 3' end of the nucleotide sequence. Typical algorithms will then identify oligomers of defined length that are unique to the gene, have a GC content within a range suitable for hybridization, and lack predicted secondary structure that may interfere with hybridization. In certain situations it may be appropriate to use pairs of oligonucleotides on a microarray or detection kit. The "pairs" will be identical, except for one nucleotide that preferably is located in the center of the sequence. The second oligonucleotide in the pair (mismatched by one) serves as a control. The number of oligonucleotide pairs may range from two to one million. The oligomers are synthesized at designated areas on a substrate using a light-directed chemical process. The substrate may be paper, nylon or other type of membrane, filter, chip, glass slide or any other suitable solid support.

In another aspect, an oligonucleotide may be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application W095/251116 (Baldeschweiler *et al.*) which is incorporated herein in its entirety by reference. In another aspect, a "gridded" array analogous to a dot (or slot) blot may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. An array, such as those described above, may be produced by hand or by using available devices (slot blot or dot

WO 02/34922

PCT/US01/42528

blot apparatus), materials (any suitable solid support), and machines (including robotic instruments), and may contain 8, 24, 96, 384, 1536, 6144 or more oligonucleotides, or any other number between two and one million which lends itself to the efficient use of commercially available instrumentation.

- 5 In order to conduct sample analysis using a microarray or detection kit, the RNA or DNA from a biological sample is made into hybridization probes. The mRNA is isolated, and cDNA is produced and used as a template to make antisense RNA (aRNA). The aRNA is amplified in the presence of fluorescent nucleotides, and labeled probes are incubated with the microarray or
- 10 detection kit so that the probe sequences hybridize to complementary oligonucleotides of the microarray or detection kit. Incubation conditions are adjusted so that hybridization occurs with precise complementary matches or with various degrees of less complementarity. After removal of nonhybridized probes, a scanner is used to determine the levels and patterns of fluorescence. The scanned images are examined to determine degree of complementarity and the relative abundance of each oligonucleotide sequence on the microarray or detection kit. The biological
- 15 samples may be obtained from any bodily fluids (such as blood, urine, saliva, phlegm, gastric juices, etc.), cultured cells, biopsies, or other tissue preparations. A detection system may be used to measure the absence, presence, and amount of hybridization for all of the distinct sequences simultaneously. This data may be used for large-scale correlation studies on the sequences, expression patterns, mutations, variants, or polymorphisms among samples.
- 20 Using such arrays, the present invention provides methods to identify the expression of the drug-metabolizing enzyme proteins/peptides of the present invention. In detail, such methods comprise incubating a test sample with one or more nucleic acid molecules and assaying for binding of the nucleic acid molecule with components within the test sample. Such assays will typically involve arrays comprising many genes, at least one of which is a gene of the
- 25 present invention and or alleles of the drug-metabolizing enzyme gene of the present invention. Figure 3 provides SNP information that has been found in the gene encoding the drug-metabolizing proteins of the present invention. SNPs, including insertion/deletion variants ("indels"), were identified at 45 different nucleotide positions. Changes in the amino acid sequence caused by these SNPs can readily be determined using the universal genetic code and the protein sequence provided in Figure 2 as a reference. Positioning of each SNP in exons,
- 30 introns, or outside the ORF can readily be determined using the DNA positions given for each SNP and the start/stop, exon, and intron coordinates given in the features.

WO 02/34922

PCT/US01/42528

Conditions for incubating a nucleic acid molecule with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the nucleic acid molecule used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or array assay
 5 formats can readily be adapted to employ the novel fragments of the Human genome disclosed herein. Examples of such assays can be found in Chard, T., *An Introduction to Radioimmunoassay and Related Techniques*, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G. R. *et al.*, *Techniques in Immunocytochemistry*, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., *Practice and*
 10 *Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1985).

The test samples of the present invention include cells, protein or membrane extracts of cells. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed.
 15 Methods for preparing nucleic acid extracts or of cells are well known in the art and can be readily be adapted in order to obtain a sample that is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention.

Specifically, the invention provides a compartmentalized kit to receive, in close
 20 confinement, one or more containers which comprises: (a) a first container comprising one of the nucleic acid molecules that can bind to a fragment of the Human genome disclosed herein; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound nucleic acid.

In detail, a compartmentalized kit includes any kit in which reagents are contained in
 25 separate containers. Such containers include small glass containers, plastic containers, strips of plastic, glass or paper, or arraying material such as silica. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will
 30 include a container which will accept the test sample, a container which contains the nucleic acid probe, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound probe. One skilled in the art will readily recognize that the previously unidentified drug-metabolizing enzyme gene of

WO 02/34922

PCT/US01/42528

the present invention can be routinely identified using the sequence information disclosed herein can be readily incorporated into one of the established kit formats which are well known in the art, particularly expression arrays.

5 Vectors/host cells

The invention also provides vectors containing the nucleic acid molecules described herein. The term "vector" refers to a vehicle, preferably a nucleic acid molecule, which can transport the nucleic acid molecules. When the vector is a nucleic acid molecule, the nucleic acid molecules are covalently linked to the vector nucleic acid. With this aspect of the invention, the vector includes a
10 plasmid, single or double stranded phage, a single or double stranded RNA or DNA viral vector, or artificial chromosome, such as a BAC, PAC, YAC, OR MAC.

A vector can be maintained in the host cell as an extrachromosomal element where it replicates and produces additional copies of the nucleic acid molecules. Alternatively, the vector may integrate into the host cell genome and produce additional copies of the nucleic acid molecules
15 when the host cell replicates.

The invention provides vectors for the maintenance (cloning vectors) or vectors for expression (expression vectors) of the nucleic acid molecules. The vectors can function in prokaryotic or eukaryotic cells or in both (shuttle vectors).

Expression vectors contain cis-acting regulatory regions that are operably linked in the
20 vector to the nucleic acid molecules such that transcription of the nucleic acid molecules is allowed in a host cell. The nucleic acid molecules can be introduced into the host cell with a separate nucleic acid molecule capable of affecting transcription. Thus, the second nucleic acid molecule may provide a trans-acting factor interacting with the cis-regulatory control region to allow transcription of the nucleic acid molecules from the vector. Alternatively, a trans-acting factor may
25 be supplied by the host cell. Finally, a trans-acting factor can be produced from the vector itself. It is understood, however, that in some embodiments, transcription and/or translation of the nucleic acid molecules can occur in a cell-free system.

The regulatory sequence to which the nucleic acid molecules described herein can be operably linked include promoters for directing mRNA transcription. These include, but are not
30 limited to, the left promoter from bacteriophage λ , the lac, TRP, and TAC promoters from *E. coli*, the early and late promoters from SV40, the CMV immediate early promoter, the adenovirus early and late promoters, and retrovirus long-terminal repeats.

WO 02/34922

PCT/US01/42528

In addition to control regions that promote transcription, expression vectors may also include regions that modulate transcription, such as repressor binding sites and enhancers. Examples include the SV40 enhancer, the cytomegalovirus immediate early enhancer, polyoma enhancer, adenovirus enhancers, and retrovirus LTR enhancers.

- 5 In addition to containing sites for transcription initiation and control, expression vectors can also contain sequences necessary for transcription termination and, in the transcribed region a ribosome binding site for translation. Other regulatory control elements for expression include initiation and termination codons as well as polyadenylation signals. The person of ordinary skill in the art would be aware of the numerous regulatory sequences that are useful in expression vectors.
- 10 Such regulatory sequences are described, for example, in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

- A variety of expression vectors can be used to express a nucleic acid molecule. Such vectors include chromosomal, episomal, and virus-derived vectors, for example vectors derived from bacterial plasmids, from bacteriophage, from yeast episomes, from yeast chromosomal elements, including yeast artificial chromosomes, from viruses such as baculoviruses, papovaviruses such as SV40, Vaccinia viruses, adenoviruses, poxviruses, pseudorabies viruses, and retroviruses. Vectors may also be derived from combinations of these sources such as those derived from plasmid and bacteriophage genetic elements, e.g. cosmids and phagemids. Appropriate
- 15 cloning and expression vectors for prokaryotic and eukaryotic hosts are described in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).
- 20

- The regulatory sequence may provide constitutive expression in one or more host cells (i.e. tissue specific) or may provide for inducible expression in one or more cell types such as by temperature, nutrient additive, or exogenous factor such as a hormone or other ligand. A variety of
- 25 vectors providing for constitutive and inducible expression in prokaryotic and eukaryotic hosts are well known to those of ordinary skill in the art.

- The nucleic acid molecules can be inserted into the vector nucleic acid by well-known methodology. Generally, the DNA sequence that will ultimately be expressed is joined to an expression vector by cleaving the DNA sequence and the expression vector with one or more
- 30 restriction enzymes and then ligating the fragments together. Procedures for restriction enzyme digestion and ligation are well known to those of ordinary skill in the art.

WO 02/34922

PCT/US01/42528

The vector containing the appropriate nucleic acid molecule can be introduced into an appropriate host cell for propagation or expression using well-known techniques. Bacterial cells include, but are not limited to, *E. coli*, *Streptomyces*, and *Salmonella typhimurium*. Eukaryotic cells include, but are not limited to, yeast, insect cells such as *Drosophila*, animal cells such as COS and

5 CHO cells, and plant cells.

As described herein, it may be desirable to express the peptide as a fusion protein. Accordingly, the invention provides fusion vectors that allow for the production of the peptides. Fusion vectors can increase the expression of a recombinant protein, increase the solubility of the recombinant protein, and aid in the purification of the protein by acting for example as a ligand for
10 affinity purification. A proteolytic cleavage site may be introduced at the junction of the fusion moiety so that the desired peptide can ultimately be separated from the fusion moiety. Proteolytic enzymes include, but are not limited to, factor Xa, thrombin, and enterokinase. Typical fusion expression vectors include pGEX (Smith *et al.*, *Gene* 67:31-40 (1988)), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-
15 transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-fusion *E. coli* expression vectors include pTet (Amann *et al.*, *Gene* 69:301-315 (1988)) and pET 11d (Studier *et al.*, *Gene Expression Technology: Methods in Enzymology* 185:60-89 (1990)).

Recombinant protein expression can be maximized in host bacteria by providing a genetic
20 background wherein the host cell has an impaired capacity to proteolytically cleave the recombinant protein. (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Alternatively, the sequence of the nucleic acid molecule of interest can be altered to provide preferential codon usage for a specific host cell, for example *E. coli*. (Wada *et al.*, *Nucleic Acids Res.* 20:2111-2118 (1992)).

25 The nucleic acid molecules can also be expressed by expression vectors that are operative in yeast. Examples of vectors for expression in yeast e.g., *S. cerevisiae* include pYepSec1 (Baldari, *et al.*, *EMBO J.* 6:229-234 (1987)), pMFa (Kurjan *et al.*, *Cell* 30:933-943(1982)), pJRY88 (Schultz *et al.*, *Gene* 54:113-123 (1987)), and pYES2 (Invitrogen Corporation, San Diego, CA).

The nucleic acid molecules can also be expressed in insect cells using, for example,
30 baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith *et al.*, *Mol. Cell Biol.* 3:2156-2165 (1983)) and the pVL series (Lucklow *et al.*, *Virology* 170:31-39 (1989)).

WO 02/34922

PCT/US01/42528

In certain embodiments of the invention, the nucleic acid molecules described herein are expressed in mammalian cells using mammalian expression vectors. Examples of mammalian expression vectors include pCDM8 (Seed, B. *Nature* 329:840(1987)) and pMT2PC (Kaufman *et al.*, *EMBO J.* 6:187-195 (1987)).

5 The expression vectors listed herein are provided by way of example only of the well-known vectors available to those of ordinary skill in the art that would be useful to express the nucleic acid molecules. The person of ordinary skill in the art would be aware of other vectors suitable for maintenance propagation or expression of the nucleic acid molecules described herein. These are found for example in Sambrook, J., Fritsch, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

10 The invention also encompasses vectors in which the nucleic acid sequences described herein are cloned into the vector in reverse orientation, but operably linked to a regulatory sequence that permits transcription of antisense RNA. Thus, an antisense transcript can be produced to all, or to a portion, of the nucleic acid molecule sequences described herein, including both coding and non-coding regions. Expression of this antisense RNA is subject to each of the parameters described above in relation to expression of the sense RNA (regulatory sequences, constitutive or inducible expression, tissue-specific expression).

15 The invention also relates to recombinant host cells containing the vectors described herein. Host cells therefore include prokaryotic cells, lower eukaryotic cells such as yeast, other eukaryotic cells such as insect cells, and higher eukaryotic cells such as mammalian cells.

20 The recombinant host cells are prepared by introducing the vector constructs described herein into the cells by techniques readily available to the person of ordinary skill in the art. These include, but are not limited to, calcium phosphate transfection, DEAE-dextran-mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, lipofection, and other techniques such as those found in Sambrook, *et al.* (*Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

25 Host cells can contain more than one vector. Thus, different nucleotide sequences can be introduced on different vectors of the same cell. Similarly, the nucleic acid molecules can be introduced either alone or with other nucleic acid molecules that are not related to the nucleic acid molecules such as those providing trans-acting factors for expression vectors. When more than one

WO 02/34922

PCT/US01/42528

vector is introduced into a cell, the vectors can be introduced independently, co-introduced or joined to the nucleic acid molecule vector.

In the case of bacteriophage and viral vectors, these can be introduced into cells as packaged or encapsulated virus by standard procedures for infection and transduction. Viral vectors can be replication-competent or replication-defective. In the case in which viral replication is defective, replication will occur in host cells providing functions that complement the defects.

Vectors generally include selectable markers that enable the selection of the subpopulation of cells that contain the recombinant vector constructs. The marker can be contained in the same vector that contains the nucleic acid molecules described herein or may be on a separate vector.

Markers include tetracycline or ampicillin-resistance genes for prokaryotic host cells and dihydrofolate reductase or neomycin resistance for eukaryotic host cells. However, any marker that provides selection for a phenotypic trait will be effective.

While the mature proteins can be produced in bacteria, yeast, mammalian cells, and other cells under the control of the appropriate regulatory sequences, cell-free transcription and translation systems can also be used to produce these proteins using RNA derived from the DNA constructs described herein.

Where secretion of the peptide is desired, appropriate secretion signals are incorporated into the vector. The signal sequence can be endogenous to the peptides or heterologous to these peptides.

Where the peptide is not secreted into the medium, the protein can be isolated from the host cell by standard disruption procedures, including freeze thaw, sonication, mechanical disruption, use of lysing agents and the like. The peptide can then be recovered and purified by well-known purification methods including ammonium sulfate precipitation, acid extraction, anion or cationic exchange chromatography, phosphocellulose chromatography, hydrophobic-interaction chromatography, affinity chromatography, hydroxylapatite chromatography, lectin chromatography, or high performance liquid chromatography.

It is also understood that depending upon the host cell in recombinant production of the peptides described herein, the peptides can have various glycosylation patterns, depending upon the cell, or maybe non-glycosylated as when produced in bacteria. In addition, the peptides may include an initial modified methionine in some cases as a result of a host-mediated process.

WO 02/34922

PCT/US01/42528

Uses of vectors and host cells

The recombinant host cells expressing the peptides described herein have a variety of uses. First, the cells are useful for producing a drug-metabolizing enzyme protein or peptide that can be further purified to produce desired amounts of drug-metabolizing enzyme protein or fragments.

5 Thus, host cells containing expression vectors are useful for peptide production.

Host cells are also useful for conducting cell-based assays involving the drug-metabolizing enzyme protein or drug-metabolizing enzyme protein fragments, such as those described above as well as other formats known in the art. Thus, a recombinant host cell expressing a native drug-metabolizing enzyme protein is useful for assaying compounds that stimulate or inhibit drug-metabolizing enzyme protein function.

10 Host cells are also useful for identifying drug-metabolizing enzyme protein mutants in which these functions are affected. If the mutants naturally occur and give rise to a pathology, host cells containing the mutations are useful to assay compounds that have a desired effect on the mutant drug-metabolizing enzyme protein (for example, stimulating or inhibiting function) which may not be indicated by their effect on the native drug-metabolizing enzyme protein.

Genetically engineered host cells can be further used to produce non-human transgenic animals. A transgenic animal is preferably a mammal, for example a rodent, such as a rat or mouse, in which one or more of the cells of the animal include a transgene. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal in one or more cell types or tissues of the transgenic animal. These animals are useful for studying the function of a drug-metabolizing enzyme protein and identifying and evaluating modulators of drug-metabolizing enzyme protein activity. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, and amphibians.

25 A transgenic animal can be produced by introducing nucleic acid into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. Any of the drug-metabolizing enzyme protein nucleotide sequences can be introduced as a transgene into the genome of a non-human animal, such as a mouse.

30 Any of the regulatory or other sequences useful in expression vectors can form part of the transgenic sequence. This includes intronic sequences and polyadenylation signals, if not already

WO 02/34922

PCT/US01/42528

included. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the drug-metabolizing enzyme protein to particular cells.

Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder *et al.*, U.S. Patent No. 4,873,191 by Wagner *et al.* and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of transgenic mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene can further be bred to other transgenic animals carrying other transgenes. A transgenic animal also includes animals in which the entire animal or tissues in the animal have been produced using the homologously recombinant host cells described herein.

In another embodiment, transgenic non-human animals can be produced which contain selected systems that allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, e.g., Lakso *et al. PNAS* 89:6232-6236 (1992). Another example of a recombinase system is the FLP recombinase system of *S. cerevisiae* (O'Gorman *et al. Science* 251:1351-1355 (1991)). If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein is required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, I. *et al. Nature* 385:810-813 (1997) and PCT International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, e.g., a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G₀ phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyst and then transferred to pseudopregnant female foster animal. The offspring born of this female foster animal will be a clone of the animal from which the cell, e.g., the somatic cell, is isolated.

WO 02/34922

PCT/US01/42528

Transgenic animals containing recombinant cells that express the peptides described herein are useful to conduct the assays described herein in an *in vivo* context. Accordingly, the various physiological factors that are present *in vivo* and that could effect substrate binding, drug-metabolizing enzyme protein activation, and signal transduction, may not be evident from *in vitro* cell-free or cell-based assays. Accordingly, it is useful to provide non-human transgenic animals to assay *in vivo* drug-metabolizing enzyme protein function, including substrate interaction, the effect of specific mutant drug-metabolizing enzyme proteins on drug-metabolizing enzyme protein function and substrate interaction, and the effect of chimeric drug-metabolizing enzyme proteins. It is also possible to assess the effect of null mutations, that is mutations that substantially or completely eliminate one or more drug-metabolizing enzyme protein functions.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the above-described modes for carrying out the invention which are obvious to those skilled in the field of molecular biology or related fields are intended to be within the scope of the following claims.

WO 02/34922

PCT/US01/42528

Claims

That which is claimed is:

1. An isolated peptide consisting of an amino acid sequence selected from the group consisting of:
 - (a) an amino acid sequence shown in SEQ ID NO:2;
 - (b) an amino acid sequence of an allelic variant of an amino acid sequence shown in SEQ ID NO:2, wherein said allelic variant is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 or 3;
 - (c) an amino acid sequence of an ortholog of an amino acid sequence shown in SEQ ID NO:2, wherein said ortholog is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 or 3;and
 - (d) a fragment of an amino acid sequence shown in SEQ ID NO:2, wherein said fragment comprises at least 10 contiguous amino acids.
2. An isolated peptide comprising an amino acid sequence selected from the group consisting of:
 - (a) an amino acid sequence shown in SEQ ID NO:2;
 - (b) an amino acid sequence of an allelic variant of an amino acid sequence shown in SEQ ID NO:2, wherein said allelic variant is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 or 3;
 - (c) an amino acid sequence of an ortholog of an amino acid sequence shown in SEQ ID NO:2, wherein said ortholog is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 or 3;and
 - (d) a fragment of an amino acid sequence shown in SEQ ID NO:2, wherein said fragment comprises at least 10 contiguous amino acids.
3. An isolated antibody that selectively binds to a peptide of claim 2.

WO 02/34922

PCT/US01/42528

4. An isolated nucleic acid molecule consisting of a nucleotide sequence selected from the group consisting of:

- (a) a nucleotide sequence that encodes an amino acid sequence shown in SEQ ID NO:2;
- (b) a nucleotide sequence that encodes of an allelic variant of an amino acid sequence shown in SEQ ID NO:2, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 or 3;
- (c) a nucleotide sequence that encodes an ortholog of an amino acid sequence shown in SEQ ID NO:2, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 or 3;
- (d) a nucleotide sequence that encodes a fragment of an amino acid sequence shown in SEQ ID NO:2, wherein said fragment comprises at least 10 contiguous amino acids; and
- (e) a nucleotide sequence that is the complement of a nucleotide sequence of (a)-(d).

5. An isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of:

- (a) a nucleotide sequence that encodes an amino acid sequence shown in SEQ ID NO:2;
- (b) a nucleotide sequence that encodes of an allelic variant of an amino acid sequence shown in SEQ ID NO:2, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 or 3;
- (c) a nucleotide sequence that encodes an ortholog of an amino acid sequence shown in SEQ ID NO:2, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 or 3;
- (d) a nucleotide sequence that encodes a fragment of an amino acid sequence shown in SEQ ID NO:2, wherein said fragment comprises at least 10 contiguous amino acids; and
- (e) a nucleotide sequence that is the complement of a nucleotide sequence of (a)-(d).

6. A gene chip comprising a nucleic acid molecule of claim 5.

WO 02/34922

PCT/US01/42528

7. A transgenic non-human animal comprising a nucleic acid molecule of claim 5.
8. A nucleic acid vector comprising a nucleic acid molecule of claim 5.
9. A host cell containing the vector of claim 8.
10. A method for producing any of the peptides of claim 1 comprising introducing a nucleotide sequence encoding any of the amino acid sequences in (a)-(d) into a host cell, and culturing the host cell under conditions in which the peptides are expressed from the nucleotide sequence.
11. A method for producing any of the peptides of claim 2 comprising introducing a nucleotide sequence encoding any of the amino acid sequences in (a)-(d) into a host cell, and culturing the host cell under conditions in which the peptides are expressed from the nucleotide sequence.
12. A method for detecting the presence of any of the peptides of claim 2 in a sample, said method comprising contacting said sample with a detection agent that specifically allows detection of the presence of the peptide in the sample and then detecting the presence of the peptide.
13. A method for detecting the presence of a nucleic acid molecule of claim 5 in a sample, said method comprising contacting the sample with an oligonucleotide that hybridizes to said nucleic acid molecule under stringent conditions and determining whether the oligonucleotide binds to said nucleic acid molecule in the sample.
14. A method for identifying a modulator of a peptide of claim 2, said method comprising contacting said peptide with an agent and determining if said agent has modulated the function or activity of said peptide.
15. The method of claim 14, wherein said agent is administered to a host cell comprising an expression vector that expresses said peptide.

WO 02/34922

PCT/US01/42528

16. A method for identifying an agent that binds to any of the peptides of claim 2, said method comprising contacting the peptide with an agent and assaying the contacted mixture to determine whether a complex is formed with the agent bound to the peptide.
17. A pharmaceutical composition comprising an agent identified by the method of claim 16 and a pharmaceutically acceptable carrier therefor.
18. A method for treating a disease or condition mediated by a human drug-metabolizing enzyme protein, said method comprising administering to a patient a pharmaceutically effective amount of an agent identified by the method of claim 16.
19. A method for identifying a modulator of the expression of a peptide of claim 2, said method comprising contacting a cell expressing said peptide with an agent, and determining if said agent has modulated the expression of said peptide.
20. An isolated human drug-metabolizing enzyme peptide having an amino acid sequence that shares at least 70% homology with an amino acid sequence shown in SEQ ID NO:2.
21. A peptide according to claim 20 that shares at least 90 percent homology with an amino acid sequence shown in SEQ ID NO:2.
22. An isolated nucleic acid molecule encoding a human drug-metabolizing enzyme peptide, said nucleic acid molecule sharing at least 80 percent homology with a nucleic acid molecule shown in SEQ ID NOS:1 or 3.
23. A nucleic acid molecule according to claim 22 that shares at least 90 percent homology with a nucleic acid molecule shown in SEQ ID NOS:1 or 3.

WO 02/34922

1/23

PCT/US01/42528

```
1  GCGGCTGCC TCCCTGCC AGGCTGAGC TCCGCTGCC ACTGCTTTC
51  CTTCTTCCG CGAGTCAGAA GCTTCGGAG GCGCCAGAG GCGGTTGGG
101 TGGGCGACCC TACGDCAGCT CCGGCGGGG GAAAGCCAC CTTCTCCGC
151 GCGCCAGGAA ACCGCGGGG TTGGGCGGT GCGAGAGCA TGGAAATTCT
201 CTGGCTGGAG ACGGCTGGG CGGGGCGCT TTACCTGCC TTCTGTCT
251 GCGTGGGCT GGGGCTGCT GAGGCCATTA AGCTGTACCT GCGGAGGCG
301 GCGTGGGCT GGGGCTGCT GCGGCCATTA AGCTGTACCT GCGGAGGCG
351 CTTTGGGCT CAGAACTTTA TTACGATGA TAACATGGAG AAGCTTGAG
401 AAATATTTGA AAAATACCT GGTGCTTCC CTTCTGCAT TGGGCTTTC
451 CAGGCTTTT TCTGTATCTA TGACCCAGC TATGCAAGA CACTTCTAG
501 CAGAACAGAT CCAAGTCCC GGTACTTGA GAAATTTCTA CTTCCACTTC
551 TTGGAAGAG ACTAGGGCT CTAGAGGAG CCAAGTGGT CCAGCATGT
601 GCGTACTAA CTCTGGAT CCATTTTAA ATCTGAAAG CATACATTGA
651 GGTGATGGT CATCTGTGA AATGATGCT GGTAAAGTG GAGAAGATT
701 GCGCACTCA GGACACAAG GTGGAGGCT ATGAGCAGT CAATCGATG
751 TCTCTGGTA TAATCATGAA ATGCGTTTC AGCAAGGAG CCAACTGCCA
801 GAGAACAGC ACCCATGAT CTTATGCAA AGCCATATT GAATCAGCA
851 AAATCATATT TCACGCTTG TACGTTTGT TGTATCAG TGACATAAT
901 TCAAACTCA GCGCTCAGG CTACGCTTC CAGAACTAA GCGGAGTGT
951 GAATCAGTAC ACAGATACA TAATCCAGG AAGAAAGAA TCCGCTCAG
1001 CTGGGCTAAA GCAGGATAAC ACTCCGAGA GGAAGTACCA GGAATTTCTG
1051 GATATTGTCC TTTCTGCCAA GGATGAAAGT GGTAGCAGT TCTCAGATAT
1101 TGATGTACAC TCTGAAGTCA GCACATTCCT GTTGGCAGG CATGACACT
1151 TGGCAGCAG CATCTCCTG ATCCTTACT GCGTGGCTCT GAACTCAGG
1201 CATCAAGAGA GATGCGGGA GGAGGTCAG GGCATCCTG GGGATGGTC
1251 TTTATCACT TGGGACAGC TGGGTGAGT GTCTACACC ACAATGTGCA
1301 TCAAGGAGC GTGCGATTG ATCTCTGAG TCCGCTCAT TTCCAGAGT
1351 CTGAGCAGC CACTTACCT CCGATGGA TGCACTTGC CTGAGGGAT
1401 CCGCTGGTT CTTAGTATT GGGGCTTCA CCACAACCT GCTGCTGTCT
1451 GGAAGAACCC AAGGCTCTT GACCCCTGA GGTCTCTCA GGAATTTCT
1501 GATCAGAGC ACCCTATGC CTACTTACC TTCTCAGTG GATCAAGGA
1551 CTGCATTGG CAGGAGTTG CCATGATTGA GTTAAAGGA ACCATTGCT
1601 TGATTTCTG CCACTTCAG GTGACTCCG ACCCCACAG GCGCTTACT
1651 TTCCCAACC ATTTATCCT CAGCCCAAG AATGGGATG ATTTGCCCT
1701 GAAGAACTC TCTGATGTT AGATCTCAG GTACATGAT TAAACGTACT
1751 TTGTTTTTG AAGTTAAAT TACAGTAAT GATCAAGCA GATAGAAAG
1801 GATCAATGTA TGGTGGAGG ATGGAGGTT GGTGGGATG GGTCTCTGT
1851 GAAGAGATCC AAAATCATT CTAGGTACAC AGTGTGTAG CTAGATCTGT
1901 TTTATATAA CTTGGGAGA TTTGAGATC TTTCTGTTA AACTTTCACT
1951 ACTATTATG CTGTATACAC CATAGACTT TCATATATT TCTGTTGTT
2001 TTAATATAG TTTGAGATT ATGCAAGTA TAAGTGCAT TATGCTCACT
2051 GTCAAAATTT CCGACACTA GAAATCATG TAGAATAAA ATTTTAAATC
2101 TCACTTCACT TAGCCGATC TCCATGCCCT GACCAATCCT ACTGCTTTTC
2151 CTAAACACAG AATAATTTG TGTGATTCT TTCAGACTTT TTCTATACA
2201 TTTATATGT AGAAATGTAG CAATGATTT GTATAGATG GATCATTCCT
2251 ATATTGTAT TGATTTTTC CACTTAATA AAATTCACCT TATTCCTTAA
2301 AAAAAAAAA AAAAAAAAA AAAAAAA
```

(SEQ ID NO:1)

FEATURES:

5'UTR: 1-189
Start Codon: 190
Stop Codon: 1720
3'UTR: 1723-2327

Figure 1

WO 02/34922 2/23 PCT/US01/42528

Homologous proteins:
Top 10 BLAST Hits

| | Score | E |
|--|-------|-------|
| gi12117369 pir IA29368 prostaglandin omega-hydroxylase (EC 1.14... | 521 | e-146 |
| gi1171661 sp P10611 CP44_RABIT CYTOCHROME P450 4A4 (CYFIVA4) (P... | 520 | e-146 |
| gi1164981 gb AAA31232.1 (J02818) cytochrome P-450p-2 [Oryctola... | 520 | e-146 |
| gi116561 emb CAA40493.1 (X57209) omega-hydroxylase cytochrome ... | 518 | e-146 |
| gi189989 pir IA34260 laurate omega-hydroxylase (EC 1.14.15.3) c... | 517 | e-145 |
| gi117167 sp P14579 CP45_RABIT CYTOCHROME P450 4A5 PRECURSOR (C... | 516 | e-145 |
| gi1203787 gb AAA41038.1 (M57718) cytochrome P-450 IVA1 [Rattus... | 510 | e-143 |
| gi189992 pir IB34160 cytochrome P450 4A7 - rabbit >gi1164985 gb... | 510 | e-143 |
| gi13738263 dbj BAA33804.1 (AB018421) cytochrome P-450 [Mus mus... | 509 | e-143 |
| gi18393238 ref NP_058695.1 cytochrome P450, subfamily IVD, pol... | 508 | e-143 |

BLAST to dbEST:

| | Score | Z |
|--|-------|-------|
| gb AW812435 AW812435 CM1-ST0181-261099-026-a02 ST0181 Homo sapi... | 1092 | 0.0 |
| gb IR56515 IR56515 yg94d06.x1 Soares infant brain INIB Homo sapie... | 769 | 0.0 |
| gb AA337301 AA337301 EST42040 Endometrial tumor Homo sapiens cD... | 640 | 0.0 |
| gb AA652746 AA652746 na65c09.s1 NCI_CGAP_Px22 Homo sapiens cDNA... | 636 | e-180 |
| gb AA863360 AA863360 ch04f03.s1 NCI_CGAP_Kid3 Homo sapiens cDNA... | 599 | e-168 |
| gb AA319338 AA319338 EST21550 Adrenal gland tumor Homo sapiens ... | 555 | e-155 |
| gb BF355963 BF355963 CM1-HT0878-060900-398-b08 HT0878 Homo sapi... | 381 | e-103 |
| gb BF445825 BF445825 nae41d04.x1 LuPSKI_sympathetic_trunk Homo ... | 365 | Se-98 |
| gb AA557324 AA557324 n181a02.s1 NCI_CGAP_Br2 Homo sapiens cDNA ... | 357 | 1e-95 |
| gb AV683266 AV683266 AV683266 GKC Homo sapiens cDNA clone GKCDQ... | 323 | 2e-85 |
| gb AW264444 AW264444 xr03d03.x1 NCI_CGAP_Brn53 Homo sapiens cDN... | 242 | Se-61 |

EXPRESSION INFORMATION FOR MODULATORY USE:
library source:
Expression information from BLAST dbEST hits:
gb|AW812435|Stomach
gb|IR56515|Soares infant brain INIB
gb|AA337301|Endometrial tumor
gb|AA652746|normal prostate
gb|AA863360|kidney
gb|AA319338|Adrenal gland tumor
gb|BF355963|head neck
gb|BF445825|LuPSKI_sympathetic_trunk
gb|AA557324|breast
gb|AV683266|hepatocellular carcinoma
gb|AW264444|brain

Expression information from PCR-based tissue screening panels:
Whole brain

Figure 1A

WO 02/34922

3/23

PCT/US01/42528

1 MEFSWLETRW ARFFYLAFVF CLALGLLQAI KLYLRQRLL RDLRPFAPF
 51 THWFLGHQXF IQDDNMEKL EIEKYPRAF PWIGPFQAF FCIYDPDYAK
 101 TLLSRTPKPS RYLQKFSPP LKGLAALOG PKWFGRRLL TPGFHFNIK
 151 AYIEVMASV KMLDKWEKI CSTQDTSVEV YEHINMSLO IINKCAPSK
 201 TNCQTNSTHD FYAKAIFELS KIFHLYSL LYHSDIIFEL SPOGYRFOKL
 251 SRVLNHYTDT IIOERKKSLO AGVKQDNTPK RKYQDPLDIV LSAKDEBGS
 301 FSDIDVHSEV STPLAGHDT LAASISWILY CLALNPEHQE RCREVRGIL
 351 GDCSSITWDO LGMSYTTMC IKETCALIPA VPSISRDLSE PLTFPDGCTL
 401 PAGITVVLIS RGLHNPRAV WRNPKVFDPL RFSQZNSDOR HPIAYLPESA
 451 GSRNCIGQEF AMIELKVITA LILLHFRVTP DPTRLPTVEN HFILKPKNGM
 501 YLHLKRLSEC

FEATURES:
Functional domains and key regions:
 [1] PDOC00001 PS00001 ASN_GLYCOSYLATION
 N-glycosylation site
 206-209 NSTH

[2] PDOC00004 PS00004 CAMP_PHOSPHO_SITE
 cAMP- and cGMP-dependent protein Kinase phosphorylation site
 Number of matches: 2
 1 265-268 RKKS
 2 505-508 KRLS

[3] PDOC00005 PS00005 PKC_PHOSPHO_SITE
 Protein kinase C phosphorylation site
 Number of matches: 4
 1 159-161 SVK
 2 278-280 TPK
 3 292-294 SAK
 4 374-376 TCR

[4] PDOC00006 PS00006 CK2_PHOSPHO_SITE
 Casein kinase II phosphorylation site
 Number of matches: 9
 1 4-7 SWLE
 2 104-107 SRTD
 3 172-175 STQD
 4 176-179 TSVE
 5 207-210 STHD
 6 292-295 SAKD
 7 300-303 SFSD
 8 302-305 SDID
 9 393-396 TTFD

[5] PDOC00008 PS00008 MYRISTYL
 N-myristoylation site
 Number of matches: 5
 1 25-30 GLLQAI
 2 298-303 GSSPSD
 3 353-358 GSSITW
 4 451-456 GSRNCI
 5 457-462 GQEFAM

[6] PDOC00081 PS00086 CYTOCHROME_P450
 Cytochrome P450 cysteine heme-iron ligand signature
 448-457 FSAGSRNCIG

WO 02/4922

4/23

PCT/US01/42528

Membrane spanning structure and domains:

| Helix | Begin | End | Score | Certainty |
|-------|-------|-----|-------|-----------|
| 1 | 12 | 32 | 1.638 | Certain |
| 2 | 76 | 96 | 1.029 | Certain |
| 3 | 316 | 336 | 1.077 | Certain |
| 4 | 395 | 415 | 1.443 | Certain |

BLAST Alignment to Top Hit:

>gi|2117369|pir|A29368 prostaglandin omega-hydroxylase (EC 1.14.15.-)
cytochrome P450 4A4 - rabbit
Length = 510

Score = 521 bits (1328), Expect = e-146
Identities = 246/493 (49%), Positives = 355/493 (71%), Gaps = 1/493 (0%)
Frame = +1

Query: 235 LAIVFCIALGLLQAIKLYLRQRLLRDLRFPAFPTFWLGHQKFIQODN-MENLZEEIE 411
+A + L L LL+A +LYL RQ LIR L+ FP PP HW LGH + Q+D +E+++ +E
Sbjct: 21 VAALLGLLELLKAAQVLYLRQMLLRALQQFPCPPFHWLLGHSREPDNGELERIQKWE 80
Query: 412 KYPRAFFPMIGPQAPFCYDQDYANTLLSRTPKSRYLQKFSPLLCKGLAALOGPKWF 591
K+P A P+N+ +A +YDFDY K +L R+DEK+ K P +G GL LGG WF
Sbjct: 81 KFGACPMWLSGNKARLLVYDFDYLVILGRSDPKAPRNYKMTFWIGYGLLELGGQTF 140
Query: 592 QHRRLLTPGFHENILKAYIEVMAHSVOMLQKWEKICSTQDTSVEVYEHINSMSLDIIMK 771
QHRR+LTP FH++ILK Y+ +H SV++MLD+WE++ S QD+S+E++H++ M+LD IMK
Sbjct: 141 QHRRMLTPAFHYDILKPYVGLMVDSVQIMLDRWEQLIS-QDSGLEIFCHVSLNTLOTIMK 199
Query: 772 CAFSKETNCQTNSTHDPYAKAIFELSKIIIFHRLYSLLYHSDIIFKLSPQCYRFQKLSAVL 951
CAFS + + Q + Y +AI +L+ ++F+R ++ + SD ++LSP+G P + ++
Sbjct: 200 CAFSYQGSVQLDRNSHSYIQAINLNNLVYTRARNVPHQSDFLYRLSFEGRLEHRAQLA 259
Query: 952 NQYTDITIQERRKSLQAGVKQDNTPKRKYDQFLDVLVLSAKDESSESSPSDIDVHSEVSTFL 1131
+++TD +IQ+RK LQ + + +++ DFLD++L AK E+GES SD O+ +EV TF+
Sbjct: 260 HEHTORVIQQRKAQLQGELEKVARRRRLDPLDVLLEFAGVZNGSSLSQDLRAEVDTFM 319
Query: 1132 LAGHDTLAASISWILYCIALNPEHQRCREVRGILGDGSSITWDQLGEMSYTTNCIKET 1311
GHDT A+ +SMI Y LA +PEHQ RCREE++G+LGGG+SITW+ L +H YTTNCIKE
Sbjct: 320 FEGHOTTASGVSWIFALATHPEHQRCRESIQGLLGDGASITWEHLDQNPYTTNCIKEA 379
Query: 1312 CRLIPAVFSISRDLSEKFLTFPDGCTLPAGITVVLISWGLHNNPAAVWKNPKVDFPLRFSQ 1491
RL P VES++R LSKP+TFPDG +LP G+ + LSI+GLH+NF VM+NP+VDFP RF+
Sbjct: 380 LRLYPVFSVTRQLSKPVTFPDGRSLPKGVILFLSIYGLHYNF-KVWQNFVDFPFRPAP 438
Query: 1492 ENSDQRHFPYAYLPFSAGSSNCIGQEFAMIELKVTIALILHFRVTFDPTREPLTFNNHIL 1671
+++ H +A+LPFS G+ANCIG++FAM BLKV +AL LL F + POPTR +L
Sbjct: 439 DSA--YHSHAPLPFSGGARNICQGFAMRELKVAVALTLRFELLPOPTRVPIPIARVVL 496
Query: 1672 KFKNGMYLHLKRL 1710
K KNG++L L+KL
Sbjct: 497 KSKNGIHLALRL 509

Hmmer search results (Pfam):

| Model | Description | Score | E-value | N |
|---------|------------------------------|-------|----------|---|
| PF00067 | Cytochrome P450 | 416.5 | 2.5e-121 | 1 |
| CE00363 | E00363 glycine_receptor_beta | 2.1 | 4.7 | 1 |

Parsed for domains:

| Model | Domain | seq-f | seq-t | hmh-f | hmh-t | score | E-value |
|---------|--------|-------|-------|-------|-------|-------|----------|
| CE00363 | 1/1 | 210 | 233 | 481 | 504 | 2.1 | 4.7 |
| PF00067 | 1/1 | 46 | 504 | 1 | 497 | 416.5 | 2.5e-121 |

FIGURE 2, page 2 of 2

WO 02/34922

5/23

PCT/US01/42528

```
1 CCAGCCTCTC TTAGGCTCCT AAATATAGTG CAAAAAGTTC CAGAGTTCCT
51 TTGTTACCCA TGAAGGACCA TGGACGGTGT CTGACAGGG GCAACTGGCC
101 CTGGAGCAGA GGAGTAACTG CATAGAACTG TCCAAGCCTC AGAGGGAGTC
151 ACACCACCAG CAAGAACCTG GGTGGAGTA GGTGAGCCAA GGGSTTCCCA
201 GGCTCTGACC CTGCCAAGAG AACTCATTAG AAGGTACCA ACCACACATA
251 CTATTCTCTG GTCTCATGAA GAACCCAGGG ACCGGACCAG GCAAGATATC
301 ACAAGCTGA AGTTTCAGCT CTGGGGCAGA GCATGGATCT GAGGTCTTTG
351 GGCCTACCA CATTGGATCA TATGAGGGCC ATCATACAAC CATCATGATT
401 TGGGGGAGGA ATAGGGCATA GAGGAATCAT ATGAAAAGCT GAAATGCCAT
451 GAGTTACCCA GAAGAAGCTG TGTAAAGCAG AGGATTCTGA GACCTGTCTA
501 AATAACAACA TCTAGTTGAA GGTGGAGTT AGGTAGGAGG TAGGGAAGTC
551 TGGGAAGAGG GGAGCTGAAA CACTTGTCTG GTGTGGCTTA ATGGAACATG
601 CAAGGGGCCA GGACGAACTT GGTCCAGATG AAGTCAACCAC CCGCTGGGGC
651 CTGCTTTTTT TTTTTTTTTT TTTTTTTTTT TGAGACGGAG TCTCACTCTG
701 TCACCAGGCT GGAGTGCAGT GGGGGGATCT GGGCTCACTG CAATCTTTGC
751 CTCTGGGGT CAAGCGATTC TCCTGCCTCA GCCTCCTGAG TAGCTGGGAT
801 TACAGGGGGG CGCCACCAGG CCCAGCTAAT TTTAGTACTG TTAGTAGAGA
851 TGGGGTTTCA CCATCTTGGC CAGGATGGTC TTGATCCCTT GACCTGGTGA
901 TCGGGCCGCC TCGGCTCCGC AAATGTCTGG GATTACAGGC GTGAGCCACC
951 TCGGGCCGCC CCGTGGAGCC TGTCTTAATC ACTTACCCGC CAATATAAAT
1001 CTGGCTCCAG AGAGTGGAGC GTAGGCTTAA GGAATTGGGG GCGGAAGGGC
1051 GCGGAAGGTG GGGGAGGGAC AGTATAGGG AGAACACCGA ATTGTAGCAG
1101 AAATTGGGTT TATTGTTTCA AGCTGTCAAT GAACACTTAA CATATGCTCTG
1151 TCTTAGCCTA AATCAATGAA TAAATGAATG AATTAATAAA TGAATGAAT
1201 GTGGGCAATG CCTATAAGA TTGCTGGGAC AGGGAGGTGG GGGGAGACAC
1251 CAGCTTGGGA AGTCAGGCTT GTTAGATCCT AGTTCACACG CTGATACGTT
1301 ACAAACTACT AAACCATCAC TTTCAAAATA TTTTACTAC ATTTCTCTGT
1351 TATCTGTACT CGAGTTTAT TATGTTCTG GCATCTAGAG TCAGCCCTTC
1401 ATGGGCAATG GACCCAGCA GCCACAGAG GCTCTGAACC CAGAGAGACA
1451 TATGCTGGT TTAATGGTCT GTCATCTTAG AATGTTAAT AAAGTTTTTA
1501 TCCGCAATTT TCATTTTGA CTGAGATTCA TAAATTAAT AGCAGGCTCT
1551 GACTGTACCT GTATAGTGA ATTAATATAT GATGGTACGC TACTGTGCAT
1601 ATCTTCCCGG TTCAGTGTTC AGTGGCTCG TATGGCAGC TTGAACAGC
1651 TCATGTGACA CGCTGGGAAT CAGGCTGGGA ATCAGTTGTA AACATTTAC
1701 CGGAACACCA CTAGGAGGCG CACAGATAA AGGAATAATG ATGTAGACCC
1751 TCCGCTACC TCTACCCTT GGAATTTTG GTAGAAATGCC AGAATGGAAA
1801 AGAAATCTC TTGCATAGCC ATTTATAAT TGTGATAAGG AAGAAAACA
1851 ATGACCTCAG CTTTAGCAT ATTTACAAT ATAAATTCAG ATCCGTGAC
1901 TGAAGACTGT TGGACTTAA AGAGGAGGCT CCAGGAGGCG AAAGCAGTT
1951 GGGCGAAGC AAGGCTGCGC GCTTTGTAA CCGGCTAGAA ATCCGACG
2001 CGGCTGCTC TCTCTCTCC AGGCTGAGC TGCCCTTCCC ACTGCTTTC
2051 CTCTTCTCCG CGAGTCAGAA GCTTGGGAG GCGCCAGAGA GCGGTGGGG
2101 GTGGGCGACC CTACGCCAGC TCCGGGCGGG AGAAAGCCCA CCGTCTCCG
2151 GCGCCCATGA AACCGCCGGC GTTGGGCGCT GCGCAGAGCC ATGGAATCT
2201 CCGGCTGGA GACCGCTGG GCGGGGCGCT TTTACCTGGC GTTGTGTTT
2251 TGCTTGGCCC TGGGGCTGCT GCAGGCAAT AAGCTGTACC TGCGGAGGCA
2301 GCGGCTGCTG CCGGACCTGC GCCCTTCCC AGCGCCCCC ACCCATGCT
2351 TCCCTGGGCA CCAGAAGGTA AATGGAAGG AAAAAAGNTA GAAAGGAGG
2401 AACAGGGGGG CCGAGGAGGA TGCGGAGAG CAGCCACGCC GCGAGAGGA
2451 CCGAGCTTTC TTCCATCCCT GGGGACCTTC GGGCTTGCAC CCGCTTTTC
2501 AGCCCGGCT GTGGCTCTTA GCATCATTTT TCCCTGCTCT GGGAATTTG
2551 TTTCCCGCAG CCCCACAGG AAGGTGACA AAGAGGAGG CTTTGGGGG
2601 TGGGAGAGAG CTATTTAAAG AACCTGAATA TGGAAAAAGA AAGCGAGCTG
2651 TAACTCAAGT CTGTCTCTCA TTGCTTACC AAGCCTTCCA CATGTGTTG
2701 TTTAAAAATA GCATGTTATT CTAATAAAT TATTAGTTGC AGAAAAATG
2751 CAAATCTAT CCAATCTGTT GGCACCTTA GTCCATTTA ACAAGAGAAA
2801 ATTJTCTTT CCTAAGATT TTGGAAGTA AGGAGCAGCC CCAGCCAGCC
2851 ACTCGAGAAA TACTGATTGA TGGAAATTT TAAAGGGAGA CTGTAGCTT
2901 TTGCTCTCT CCGTTTTTA AATCCACTCC CACCCCTAAT TAAGGTTTTT
2951 ATTCATCAA CCGACTCTGA GTGGCAATT TGTGATAGGT ACTAAGATTA
3001 CAAAGAGAAG CTAAGTCCCT CCGCTGACC ACCCAAGTCA GGTGCAGACT
3051 TAGGCCACAG AGAGAAATG AAAATTTAAG GCAATGGGTO CTTTACTAGA
3101 GCGCTAGAGA CAAGGAATA TCTGTGGAG GAAAGTATAC ATCTCGGCT
3151 AGAGAGGAA GGAAGTCTG TGAAGGGCTG AGCAGAGTCT TAAAGGATGG
3201 TTGGGTGGTG TGGGAAGGC ATTCAGCAG AGCTACTACA CGATCCCTTG
3251 GTTCCCCAC TTTCTAGTCT TTTTATATA AAGCAACCAC TTTCAACTCT
```

FIGURE 3, page 1 of 19

WO 02/34922

6/23

PCT/US01/42528

3301 TTTATCGGT TCTTCTGGTA TTTAAATACI TATTTGTAAA ATAGTATTAC
3351 CATATGCA7 CTATTAAIT7 AATAAGTT7A GACATCTGCT GTGGTTTAGA
3401 TATGGTTTGT TCGTCCCCAC CAGGCTCAT GTTGAATTT GATCCCAAT
3451 GTTGGAGGTG GGTCTGATG GGAGATCTT GGGTCATTGG CATGGATCC
3501 TCATGAATGT CTGGTGCAG CTGCTCCTT CATAAGTTCY CACTCTCTTA
3551 GTCCCTCTTC AACCCCCAGA ACTGATTGT7 GAAAAGAGCC TCCACCTCC
3601 TCCCTCTCT CTTCCTGTCT CTCACCATGT GGTCTCTGCA CACAACCTGCT
3651 CCTGTTCACT TCCACTATGA GTGGAAGCAG TCTGAGATCC TCCGAGATG
3701 CAGATGCCAA TGCCATGCTT CTGTACAGC CTGCAGAT7 GTACCCCAA
3751 TAATCTCTT7 TGGAATGAC CCAGCTCAG GTATTCTT7 ACAGCAACAC
3801 AAATGTACTA AGACAACATC CACTATGAA CTCTTTATG ACAGGCAATC
3851 ACTTACACTT CATATTCCAC TGTCCAGTA ACTATATAGT ATTGTATT7T
3901 TTAATAGAA7 AAACCTCTAT TGTATTAT7 TTTATTATGC AAATGT7AT7
3951 TACTGCTGAT CTAAATGGTC CTCTTTCATT TTAATTCCTT TTCTCA7AGA
4001 ACTTTTCC CACCCCCACA GTATTGNNNN NNNNNNNNNN NNNNNNNNNN
4051 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
4101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
4151 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
4201 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
4251 CTATGCTGTC TGTGTTT7A ATTGAATGT TTTGGCATCC TTGCAAAA
4301 TCATTTGACC ATAAATGTCA AGGTCTATT CTGAGTCTTC AATTCTAATC
4351 CATGATCTA TATGCTATC CTAACCTATG GACACAGAGA GTAGAAGGAT
4401 GGTACCAAAA GGCTGGGAG GATAGAGGG AGCTGGGGGA GGAGGTAGGG
4451 AAGGTAAATG GGTACAAA7 AAATAGAA7G AATGAATAC ACCTACTATT
4501 TGATAGCTA GCGGGTGGC TATAGTCAAT AATAACTGTA CACTTTAAA
4551 TAAAGAGTGT AATAGGAT7G TTTGCAATC AATGGATAAA TGCTTGAGGG
4601 GATGGGTACC CCATTTCTCA TGATGTGCTT ATTTACATT GCATGCTGT
4651 ATCAAAACA TCTCAATTAC TCCATAAATA TATACACCTA CTATGATCC
4701 ACAAGTATTA AAAATTATAA ATAAATAAAT TATATAGCTA TCCATATGCT
4751 AGTACCACAC TGCCCTACTG TTGCTTGT7A GTAACTTTG AAATCAGGAA
4801 GTATGAGTCC CCGGCACTT7 GGTATTTTCC AAGATTATT TGGCTGTTG
4851 GAATCCTTGA TTCTATAG7A AATTTTAGAC TCAGCTATTC AATTTCTACA
4901 AGGAAACAG CTAGGGTCT GCTTGGGATT GCACTGAATC TGTAGATCAG
4951 TTTGGGATT7 ATTGGCATCT TAAGATAT7T AGGTCTCTG ATCCATGAAC
5001 ACAGAAAGCC TTTCCGTT7A GTTAGGTCAT CTTTAATTTT TTTGTTGTT
5051 TTTTCTGTT TTTGAGACA GAGTCTGCT CTGTCCCA7C GGCTGGAGTG
5101 CAGTGAGGCA ATCTGGGCTC ACTGCAACCT CCGCTCTCG GATTCAAGCG
5151 ATTCTCTGC CTGAGCTCC CAGCAGCTG GGACTACAGG CACATGCCAC
5201 CACACCAACT AATTTTGT7A TTTTCAGTAG AGAGGGGGTT TCACCATATT
5251 GGCCAGGCTA GTCTGAACT CCGACCTCG TGATCCACCC GCTTCACCT
5301 CCAAGTGC TGGGATTACA GGCGTAGCC AGCACTCCCG GCTTCTTTA
5351 ATTTTCTTA ACGATGTTT TGTATTTTC AAAGTATACA TCTTGCAATT
5401 CTTTGT7TA ATTTATTGT TTTGTTCTT TTAATTTCAT TCCAGACTAT
5451 TTATTGCATT CATAGT7TT TAGAGTCCAC ATTCCCTCT GACTGTCACT
5501 AAGTTT7TT TTTCTGT7T TTAGAGGTT TCTATCAGAA TTTTGAGAT
5551 CAGAGATGAC GGACATGTCA AACTGTCTAA TATTACCAAC CTTCCCACT
5601 TATCAGATCA GGATCCTTT GGTGATTAC CATGCAGGGA AATCTAGTAT
5651 CTAAGGCTCA AAAGGTGATA CTGTTTACA TAGGCA7TA CATTTTATTG
5701 CTACATAATA ACTACATATT TATGGAGTAC CTGTATATT TGTATAGTG
5751 CATACAATGT GCAGTATCA AATCAGGGTG TTAGGGTAT TCATCACTTC
5801 TAACATTAT TATTATT7G TGT7TGCAC ATTTCAAGTC TCTTCAAGCT
5851 CTTCAAAAT ATTCAATACA TTAATT7TA CAGTGGCTAT GAACACTGGA
5901 ACTTATTCT TCTATCTAAA GACAGTAA7A TTTTAA7AT AGTCATAAGG
5951 TACAGAGG ATAAAGTGTG TATAGGGA7A ATTCCCTACA AGATGAGAA7
6001 TTCA7TCTT ACTCTTAGTA ATACAGTCT TCAACATGC CAAGGATATT
6051 CTTCCCTTG AGCTTTGAAC ATGCACTCT GTGGTTATAT TGCTCTCCCT
6101 GCAAA7TATT CCTAAAAGAG GCTTCCCTG ACCATT7CAG CTAAAATAGC
6151 AACTCTAGTA CTCTCTATCT CCACCCCTAT TATTATTATC TTGGCCCTTA
6201 TCACTCTCTG ACACTATACT GTATACTCT TTGCTTGTTC GTTTATTATC
6251 CACCACTAAC TACAATATAA AATCTGTGAG AGGTAGGATC TTTGTTTGC
6301 ACTATAMCC TAGTGCATGG TACAGTTCTT GGTGCATAAT AGGTGCTCAA
6351 TAAATCCTTT GTTGAATGCA TAAATATAT AGGTGCTGAG AAATTTATT
6401 TATTCAAGA TCAATTACT GCATAGATA GGCCAGGTGG TTTGACATT
6451 ATTCAATAGC CAACATATGG GACCTAGGAT GTACATATOC AAGTGTGTG
6501 GTGTATGCT GTGTGATCT GCATGTGAC TTGGATGTAC TGCAAGAAC
6551 ATCTATG7AG CTAAGTAGTA TAAAGCACT GGGCTCCAGA GTTAAACTGG

FIGURE 3, page 2 of 19

WO 02/34922

7/23

PCT/US01/42528

6601 AGTTTGAATC CTCATTAGTG GTTGCCAGCT GTACACACTT GGGCAGATCA
6651 TTTAACCTAG TCTGYAGGGC TCAATTTCTT CATCTCTAAA GTAGGGATTC
6701 TAATCATATC TACTTCATAG GGTTCCTGAT GTAAATATTA AATACATAG
6751 AACATGGAAA GCATTTAGCA GCACCTAGTT CATAGCAGTG CTTGATAAAT
6801 GTTCGCTGTT GCTATTTGGG GGCACATGCT ATTTCTGAA CATTTCTGAA
6851 CAATGTTTAC TAAATATATG TAGTACCGGT TTTCAAGTGT ATTTAGATGC
6901 TTCTCTGGGG ATGAAGAAAT ATAAATTAAT TATAGTACAG TATTACAAAC
6951 AGTTTCTGT CTTTTTGTG TAGTCAGGAG TTACAAAAG TATAATGAAA
7001 TACTTTCATA TGCTGGGGT GTTTATGAAA ATTTTTTACC TAAACAAACA
7051 ATTGTCATAT TAGTTTACAA TATTCAATG GGCAAAAGGC TTGCTTCTCT
7101 TATATTTCTC TGTATCTCTA CCACCTGGTA CGTGTGATAG ACAATAAATA
7151 CTGTGTGTT TATTGTTTGT AAATGAATAA ATGAAAAAAT ATTCACTTTG
7201 TTGAAAACCA CTACTCTGGA TAGTCAGTGG GTGCTTATCA CTGGCTTGAT
7251 TATGGCAACA TTACAAAAA AGTGCAGTAT TTTAGAAACT AGGTTTCAAG
7301 ACTCTCAACC TTTCACTGGC CTTGAATAT CCAGAGAAC CTTTATGGGT
7351 TAAATTTGCT AAATGATAAC AGAGAAAAAT GGGAGCCAGA GTTGTCCACC
7401 TCTCCAGAGG ATGAGAGCAA ACAATCTGCT AGCAGATACC GTGTGATTGG
7451 TCACACGAGG AAAAATCTGG CAGCCTTAAG ATTACTTTGC AGCGGGGAC
7501 TCCCAACATC ATGCTCAAGT GTGTAGATGG GCACACCAAA ACACACACAT
7551 GCAGGTGGCC TCCACTTTAC ACAAGAGCA AATGATAATG AATCTTGTTC
7601 TCAGTGATTT AGAGAAACAA TTTAAGTGAG CCATTACTCA TCTGCTCTA
7651 AAAGCAAAAA CTCTCTCTCT GGTGGTAGTA TTTGCACTCT CATTTGAAA
7701 TGTGGAAAGC TGAAGTTTGT GTATTGAGT TTGCTTTAAG ATTCAACAT
7751 CTGTCTAAAT GCACTTCTG TTGTTGGGG GAGAAATTTG ATTTCTTTA
7801 TAGATAGAGT TGGCAATTTT TTAGAGAGAA GCATTTACTG CTAAGTCATG
7851 AGAATTAATC ACTGCTGCAT AATTAGAGAG AGGAACAGGA AGAAGAAATG
7901 GTAGGCTGGA TGTAGGTCA TCCCTCATTT AGTAAGTGT AGTTTCCAC
7951 ATAGGAAATA CTCTTTTGA GCTTCCAGT CCCACTCCAA TCTGAGTGTG
8001 TGATGTTGGC AAGTGAGGCA GAGAGTGGA CTGGCTCAC CCTCTATTGG
8051 GACAGAGCTT CACAGTAAAT GTCAATCAAC AGTGACTTGG TCTGGGGTA
8101 CAGGATATAT TAATATTGAG AGATAAATA CACTAATCTT GTTGAAGAA
8151 TTATCCCCCA AGCTTAGAAG TCCCAAGAA AGCATGTTAT GTCACCTCCA
8201 GAAAGTCTC AGGCTCTCT GCTTGTGCA CCTTATCAGG TCTGAACTC
8251 AGCTTGTGTC TATAAGAGGG GACAGGTCCA GCTTGGCTGG CTAATTAATT
8301 TTACTTTTCT CACTGCAGTT TATTGAGAT GATAACATGG AGAAGCTGA
8351 GGAATATTAT GAAATAAOC CTGCTGCTCT CCCTTCTGG ATTTGGGCT
8401 TTCAGGCAAT TTTCTATATC TATGACCCAG ACTATGCAAA CACACTCTC
8451 AGCAGAACAG GTAAAGAGAG GGGGAAAGCT CTGGGACCTA TTCTCTCTAG
8501 AAGTGAAATG CATAAAAACC ATAGGCAAGA TTCCAAAGCA AAGATTGGTT
8551 TGGGGCTTAT AAGAGACACA GCAGCAAGTA TGGGGAGGTG ACAGGTCTCC
8601 TACCAATACT GAAGGGGAT CCAATATCTT CCCAGTCTCC TTGCTTGTTC
8651 CAGGTATGCA TGGGCACGTT GAAGTCGGTA TAACCTAAAG CCTAGCTGGC
8701 ATTACAGAGC TTGCCAGGCA AGGCTTCCCT TGGCTCTGT GGGTTTATG
8751 ACTTCAGTGT CAGCAACACT TCCCACTCTT ACCCTTGGTC TCGAGCATAA
8801 GTCTCAAGAG GGTGGGAAAT CAGCAGTAAC TCTAOCCTG CTGGTTCACT
8851 ATGAAAGCCT GAATGCTAGA TCATTAAATT ACCCATCAGA CCTCTTGATN
8901 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8951 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
9001 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
9051 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
9101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
9151 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
9201 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
9251 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
9301 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
9351 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
9401 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
9451 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
9501 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
9551 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
9601 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
9651 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
9701 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
9751 TGACTCTGCA GATCCCAAGT CCACTACCT CGAGAAATTC TCACCTCCAC
9801 TTCTTGGTAT GTATGTGCAA ATGAGAGGTA TAACCCACTC TCATTCAAG
9851 TCCCTTTTCC ATAGTAGAGC ATGCCAAGA AACTGAATC TGAATTCANA

FIGURE 3, page 3 of 19

WO 02/34922

8/23

PCT/US01/42528

9901 AGCACAAAGA GTGCAAGSTA GAGCTATACT GAAOGTTATC TAGGGGAAAG
9951 ATTGAAGGGG AGCTCTAAGG TCAACACACC ACCACTTCCC AGAAAGCTTC
10001 TTGATCCGTT TCTCTCCAC AAAGTCTTAT TCTCAAGGCA GCAGATACAT
10051 GAACTGTGCC OCTCTCTCTT TAAAACTACA GCCTTGGGCA GGCACAGTGA
10101 CTCATGCGATG TAATCCAGC ACTTGGGAG GCCAAGGTGG GAGGATCACT
10151 TGAGGTCAAG ATTTCAAGAC CAGCTGGGCC AACATGGTGA AATCCCATCT
10201 CTACTAAAAA TACAAAAATT AGCCAGGCAT GGTAGCATGT AGGCCTGTAG
10251 TCCCACTACT TGGGAGGCTG AGACATGAGA ATCGCTTGAA CCTAGGAGGT
10301 GAGGCTTGGC GTGAGCTCAG ATTGTGCCAC TGCACCTCCAG ACTAGGTGAC
10351 AGAGCAAAAC TCTGTCCGCA GCGCCCAACA ACAAAAAA AACTAGCCAA
10401 ACTGCAGTCT CACCATCCCT ATTCTGTGTT TCTTTATCCT TCTCTGCTT
10451 TCTTGGATGT TTTCTTTCT TTTTGGAGTT OCTTTATTTT CACATGCGAG
10501 TCAGTAAAAA TTTGCTTAG AGTTTGGCAA TATTCTGTCA GCAGATAAAC
10551 TAAGCTCTTT AATTACATAA TTGGTATTTA TGTAAACAA GACATGAATG
10601 AAAGAAAGA ATATAGGCTT GTATTAGGAA CCACCTAANT TTGAATCTTG
10651 CCCCCTCCTG CATTGACTAG TTAATATGA TCTTGGGAA GTCATTAA
10701 CTCTCCCTAT CTCAGTTTCC TCATCTTTGA CAATAAGGAT GAGACTCACA
10751 TTGCTGGGCT GTTATGAGCA TTAATGAAA TACATATTTT TAGCACTACA
10801 TGTAATGGCC ACCATGTAT GAGTGACAGA TCATGCTCA TGAGCCTGGA
10851 ATGTTGTAAG CATTCAATGA ATGGTATCAA TTATGTATTA ATAACTTTA
10901 AAGTCTTTT AAAGCCAAAT OCTAATGACC AGTCTGGCAA TAGAAGATTG
10951 TGAGCAATTA GCTTGGTAA GTATTCCAC ATAGTATCAT TCATAGACCT
11001 GGGCTCAAGG AGSAAATATC AGGGGACAGA GTGGACACTC TTCTCTCTTT
11051 CCTGTGAAT TTATGTTTAT CATATAGTTT ATGGATTGGT TTGAGTGGCA
11101 AAGGAATTC AATTGCTCTGT TACTAGTGTG AGCTAGGGAG TAGGTTGGCT
11151 ACCTTATGTA TTCACTTTCA GTTAACCTCC ACAGCAACAC AGGGAAAAAG
11201 GTATTTAGTA TCATAGTTCA TTATTGAGAA AAGTAAACCT CAGGAAGATT
11251 GAGTCACTTA TTCACTTACT ACATAGGTAG TAACTGGTGA TTTGAGGATT
11301 AGCGTGTAA TCTTATAAGG CTTTGAATTT TATTAGACTT TGAACCTGTT
11351 TCTCACAATA TTAATATACAT CCATCCAGA GGTAAAGCTC TAAATTCACC
11401 TTCTCTATT AAATTGCATT GCACATTAAT ACGACTACTA CTTTGTACTT
11451 CCACTGTTGC ATGACTGCTC GTGGGTCAAT GTTACTCCAC GCTGCCTGAG
11501 TTCTCTATCT ATCTTCAAT TCATCTAATT AAATGGCATA AGGTTTCTG
11551 CTTTATTTT CTCAAGSAA AGGACTAGCG GCTCTAGAAG GACCCAAAGT
11601 GTTCCAGCAT CGTGGCTTAC TAACTCTGG ATTCCATTTT AACATCCTGA
11651 AAGCATACAT TGAGGTGATG OCTCATTTCTG TCAAAATGAT GCTGCTAAGT
11701 AAAGGGGGAA AGTGTCTCTGT GCATTCGGA ATGCTCCAG CAATGACAG
11751 TATTAGGTAT GTGTTTGTG GGCCATGAAA ATAAAAAATC AGTTTCTAAA
11801 AATTTAACCA ATGTACAGT ACTTATTGAA CAATAGGTGT CTGTAAAAAA
11851 TTTGTTATGT TCTTTGAGTG ATAAATTAA TAAAAAGATC TGGTCTCTG
11901 TCTTAGATAT ATTTTGAGAT TTTATGGCAG CAACCAAGT ACCAAATGGT
11951 GATAGTTAGA TAGTAAGTGC TGTAGATGTG TTTATGGAG GCGGGTCTG
12001 TACAAACCTA CCCCAGGTC TGAGGAAACT GAGAGGCTGA AGAAAAAGGC
12051 TGACAGTTTC TAAAAAGAA ACATTCATA GAGGCTTTCA AACAAAAACC
12101 ATNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
12151 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
12201 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
12251 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
12301 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
12351 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
12401 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
12451 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
12501 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
12551 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
12601 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
12651 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
12701 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
12751 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
12801 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
12851 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
12901 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
12951 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
13001 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
13051 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
13101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
13151 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN

FIGURE 3, page 4 of 19

WO 02/34922

9/23

PCT/US01/42528

13201 NNNGNNNNNN NNNGNNNNNN NNNGNNNNNN NNNGNNNNNN NNNGNNNNNN
13251 NNNGNNNNNN NNNGNNNNNN NNNGNNNNNN NNNGNNNNNN NNNGNNNNNN
13301 NNNGNNNNNN NNNGNNNNNN NNNGNNNNNN NNNGNNNNNN NNNGNNNNNN
13351 NNNGNNNNNN NNNGNNNNNN NNNGNNNNNN NNNGNNNNNN NNNGNNNNNN
13401 NNNGNNNNNN NNNGNNNNNN NNNGNNNNNN NNNGNNNNNN NNNGNNNNNN
13451 NNNGNNNNNN NNNGNNNNNN NNNGNNNNNN NNNGNNNNNN NNNGNNNNNN
13501 NNNGNNNNNN NNNGNNNNNN NNNGNNNNNN NNNGNNNNNN NNNGNNNNNN
13551 NNNGNNNNNN NNNGNNNNNN NNNGNNNNNN NNNGNNNNNN NNNGNNNNNN
13601 NNNGNNNNNN NNNGNNNNNN NNNGNNNNNN NNNGNNNNNN NNNGNNNNNN
13651 NNNGNNNNNN NNNGNNNNNN NNNGNNNNNN NNNGNNNNNN NNNGNNNNNN
13701 NNNGNNNNNN NNNGNNNNNN NNNGNNNNNN NNNGNNNNNN NNNGNNNNNN
13751 NNNGNNNNNN NNNGNNNNNN NNNGNNNNNN NNNGNNNNNN NNNGNNNNNN
13801 ACAGCTCCTC TCCACACTTG CTCTGTTTCT CACTTTTGAA TCCAAACGTT
13851 TTTGAAAATG TTCTGAGTTT ATTTTAAAAT GTGGCTATGG TGGTTGAGAG
13901 CAGTGGCAGG GTACCTAGCA AGTTTGGAAI TGAAGTTGGA GGAAGCCCTG
13951 GGGTAAACCC CTGTAAATTA TGGGTCTTGT GTCAATGATT GCTTTAATGG
14001 AACTCTGGTC TGTTTGAAAG CAGAGTTATG GTAAATATG AAAAGCCGCA
14051 GATCTTTAAC TCAGCCATT? ACCATATATG CAGTTTCTC CATGCTCCTT
14101 CTCACTCCGC TGGGTGTATT TTTCCTTCC TCGTGCCCTG TGTAAAGCACA
14151 TGGCTTATTT ACTCATGTGA TCTTTGGTTC CTGCTGGCTC AGGGTTGTCT
14201 CCATTAGATC ATAAAAACAG GGCACGGCAG GAGCCTTCAA ATGAAGGCAA
14251 TTTGGTCATG GTGGTGGTGA TGTGTTGGT CTTGAOCTCC TGTGCCAGGA
14301 TAAAGTGGAG AAGATTGCA GCACCTAGGA CACAGCCTG CAGCTCTATG
14351 AGCACATCAA CTCGATGCT CTGATATAA TCATGAATG CGCTTTGAGC
14401 AAGGAGACCA ACTGCCAGAC AAACAGGTCA GTGGTGGGAG AGCAAAAAG
14451 ATATTTCTTC ACATTTCTA AGTTGTTTAT TAACACATTA TCCCACTTT
14501 CTCTTCTAGC ACCCATGATC CTTATGCAAA AGCCATATTT GAACCTCAGCA
14551 AATTCATATT TCACCGCTTG TACAGTTTGT TGTATCACAG TGACATAAAT
14601 TTCAAACTCA GCGCTCAGGG CTACCGCTTC CAGAAGTTAA GCGAGTGTI
14651 GAATCAGTAC ACAGGTATTT GTTGGCTTTC GGTGGCCAC GTCCATACGC
14701 TGCCATGATT GTACTGTGTC TGTCTAGAGG GATAAACCTT AATATGACAA
14751 GAGAAAGAA CTTTGTTATT AATGGAGCTT TTATATAGAC ACTGCTCCAA
14801 AGAAATTTGA CTTGAGTCTT TTATAAGACT TTGCTTCAAC CATAGCAGTA
14851 TTATCAGAA TTTTATATAT ATATATATAC ACTATTTTAA TTATGGACAA
14901 TTATTTTAA TACAAATATA AGTAGGCACT TAAGAGTTCC AGACATACAT
14951 GGAACATGSC TTTTGCACA GCGATTGCAG TAATATTAAT GHCAAGCTAA
15001 AAACATTCAI GCACATAGG AATGGAGAT GGAACAGAT AAACATGGAC
15051 ATGCACCGCA AAGAATATTG ATTCAAAAC AGTTTATGCA AGCATAAACA
15101 CAAAAGTTGA AATAGATTAA GCTTTTAAAG CAATTCACCA TTACTTGTCA
15151 TGAATGCCAT AATGGAGAAI ACTTATCAAG CAGTGAATTA ATCCTTCATC
15201 AGCTTCACCA CTTACTAGCA GTTACTAGTA AGTTACTTAC TGCTTTGTTT
15251 CAGTGTCAIC TATAAAATGG AGATTAAAAA AGAACCTATC TCATACATTI
15301 GTTGTACGA TGAGTGGGTT AATATATATA AAGCATTAG GACAGTGCCT
15351 GGCACGTAAI AGATGTTAAA TGTAAAGTAT AGTTATGTCA AATGTCTTTG
15401 CTTCCAGGAA TTTTGCAAGA CACACCAACA TATGCACACT TACACATACA
15451 TATATGCATA CATGCACATA GATATTATTA AGAGGACACT CAGAGAGCA
15501 GGTATATAAC AATTTAAGGC ATAAATGGGC ATTATATAA GCAGCAGTTC
15551 CCAAGTCTTT CTGCATCATT GCACACACAG AAAATGTTAA TGTTTTGTG
15601 CTTTATTTGA GTAAACAGGA ATGGATTGG GGGAGGCTAT ACAGAACTTT
15651 GTAAAAAAA ATCTTTACTT TTTAAATATT ATACAATTAT GATGAAAAAG
15701 CAAAATGCAA AGTGTAGGG AAAATATTAA ATGTTAAAT TATTCAAAC
15751 TTAACACCTT TTCAATTTT TTTTITTTT TTTTITGAGA TGGAGTCTCT
15801 ATCACTCAGG CTGGAGCGCA GTGGTGTGAT CTCAGCTCAC TACAACCTCC
15851 ACCTCCAGG TTCCAGCAAT TCTCTACCT CAGCCTCTG AGTAGCTGGG
15901 ATTACAGGCA CTGCCACCAC ACCTGGCTAA TTTTITTTAA TGTTTTATT
15951 TTATTTAGTC AAATATATCA ATATTTTATT TTATTTGATC TGGATTTTAA
16001 GTAATCACAA AAAGCCATTC TCTATTCCAG GCTTTCTCAA CCGTCAGCAC
16051 TAATGGCTTC TTAGATTAGA TAAGTCTCTG TTGTCAAGAT GTGTGCATTG
16101 TAGGATGTTT AGCTACATCC CTGACATCTA CCCACTOGAT GTAGTAGAGC
16151 TCTGATAGTT ATAGCAACCA TAAATAACTC CAGACATTAT TGAATGTTC
16201 CAGGGCCGCC AGTTGAGAAC CACTGCCCTG TACCCAGGTT GTAGAGAAAA
16251 TTATTTATGT TTTCTGTAG TACTGTGATA ATTTCATTAT TTTTATATT
16301 AAATCAGAGA TCTAACTCC ATTTAGAAAT TATTCTTATA TATGGTGTGA
16351 GGTATTGATC TAATTTTTC AAATGTTTAT CCAAGTTGTC CATCAOCCAT
16401 ATTTAAAGT TTATCTTTTC AAGTGATTG AGATAACCAT CACATTCTAA
16451 ACGGATACAT GTACTGATAT CTGTTTGGGA TAACAGTATA TTTGGATGTT

FIGURE 3, page 5 of 19

WO 02/4922

10/23

PCT/US01/42528

16501 CTGGTGATT CCATTGATCT ATCTACCAAT GTACCAGAA CACACTGTTT
16551 TAATTAAGGA GATTTTGTGG CTTTTTCAA CATTAAATGA CCTTATTTT
16601 AGAAAAGTTT TAGGTTTGCA GAAAAATCA GCAGAAAGTA CAGAGAGTTC
16651 TCATATTACC CATGTAACAA ACCTGTACAT GTACCCCTGT ATCTAAATTA
16701 AAGTTGAAA TTTTTAAAT AGTAAATAAA TATTACCTCT GTTCCAATAT
16751 TTGTTTGTG TTTTTCTC TCAGCTCCTT CAATTAATAA TATATTGGCA
16801 TTCTTTTGGC TGCTTCTAT TTCAATCCAT TTTATTAAAT AACTTTTGGC
16851 TGAAGATAAA ATATTAGACT GAGGAAGAAA AGAATAATTG GTCACTTGCA
16901 TCTAAACTTC AAATCATCTT AATTTTATG CCACATACAT GATGGAACT
16951 ATGTTTTTA TTTGTGTTGT TTATCTTGG AGCTTTAATC AAAAGTCCCT
17001 TTGATGAGAA AATAAACCAT CTGTGAAAAT TAGATCTATT TAAAGCTCTG
17051 GAAATCAGGC AAGATTGAA GCTATTCAT AACCATGGCT TGCTTTATAA
17101 TTTATTGAC TTTGCCATCA CTTGGTAAAT TGGAAACTAT TTTCTACCC
17151 AGATACAAAT ATCCAGGAAA GAAAGAAATC CTTCCAGGCT GGGGTAAAGC
17201 AGGATAACAC TCCGAAGAGG AAGTACCAGG ATTTCTGGA TATTGTCTT
17251 TCTGCCAAGG TAAATCTTCT AATTTCTAA GCTGCTCAA GTGACCGTT
17301 AATTAATGTA GTAGGTGGT AAGTGGGAA GGGATGGGA GACAGGAATA
17351 AAACCGATTG ACTAAATTTA ACTGTACTTT GAATTGATGA GCAGCTTCAT
17401 GCATTTGAG ACAAGAGAG AATTGTGCA CTGTGCTGT AGAGGAGGCT
17451 TAGTAAGAC TAAACGAAAG ATTTGACAA ATTTGAGGAT TGTCAATGG
17501 ATACATGCA TTTAGGCAAT CATGAATAA TGGTCACATG GATAAACGTA
17551 AAAATTAATG TGAATAGGT CTGGGAATC TGGGAGTTG AAGAGAAAT
17601 CTAGGGCTGT TTGATCGAGG GCGCTTGTG CAAGGCTGTG TTTCTTATC
17651 TAACCTTGGT TCTCTTAT GCTTTGGGA GAATATGGT TATACCAAT
17701 ATTTGTTGAA CTGAATTAAT ATTTAAACC CTATTTAAG CTCTGATTT
17751 TCCCTCAA TCAATATTGT GGTGTATCT CCAACATAT ATAACTGGC
17801 ATTTATTTA AAATATTGT ATTTACTTT CTAGGATGAA AGTGGTAGCA
17851 GCTTCTCAGA TATTGATGA CACTCTGAAG TGAGCAATC CCGTTGGCA
17901 GGACATGACA CCTGGCAGC AAGCATCTCC TGGATCCTT ACTGCTGGC
17951 TCTGAACCTT GAGCATCAAG AGAGATGCCG GGGGAGGTC AGGGCATCC
18001 TGGGGGATGG GTCTTCTATC ACTTGGTAAG ATCTGCAACC CTAATTTTC
18051 CTGCTAGTTT TCCCTGTGAG ATTTGCTTT ATTTTGTGG CTGGTACCTT
18101 AGTGACCTTA GTGCTCAGG ATATGCTAG GTGAACAGA AGAAGTAGGC
18151 TACTTTCTG TCTTTCTAA AGAGAGCTCC AAATATTCT CTTGTCTTT
18201 AGCAAAAAA AAAAGTTTA TTTATCCATA AATTGCTGT CATTTGTTT
18251 CTAATCAATG GTGTGTGAAA TGCTTATTT CTTATTTCA CTTGGCTCT
18301 GATGCATTGG AAATGAGSAC TTGATCCCTG GGCTGGCCTC TAGAATTTAA
18351 ACAATAGGGT OCAAGTGGAG CTCTCTTCT GAGAGAGCTG AATGATAGC
18401 TGCAATATTT AAGGCTCAT TTAGACATCT CCGAGCGCT TGTCAACAA
18451 TTTATTCTC AGGATTGAT TTAGACTTCA GACATAATAT TCGATGATAT
18501 ATACTATAGT TAAGTTTAGC AAATATGGAC TGAGGACAT TTAATACTG
18551 AGACTTTT TATGACTACA ATTTATTTG GGCCTGTCT TCGGTGAGCT
18601 AATGTTCTAA TACAGGAGAC AGGAGACAGA CCTCCAAAT GCAGTGTAGC
18651 ATAATGAGGG CAATGATAGA GATATGTCT GGCTAACACA AAGACATAGA
18701 AGACAGGTAC CTACCTTGGC ATGGGAGCTC AAGGAGACTT CTTTGACATT
18751 TAGCTGACT GCAGGATAAG TAGGAGTTAG CAGGTGGAA ACTGTCTCT
18801 CTATCTTGT AGACTTTAAG CATATACTG TGTAAATAA GCCCAGGTTA
18851 TGCTGTTTGC AAAGATAAAA TGTGTTCTG ACATAACTT GGTCAAGGG
18901 ACAGAAAGAC AGAATCTTA AGGACAATC AGCAGCAGC CAGATAAAAA
18951 ACACCATATT TCATATGCAA AAGTCAATC AATGAAACA TTTGTAATAC
19001 CAATTTGAC ATTATAAAG TATATCAGAG ATCTCATTT ATAGGAAAT
19051 AGAAGGCTT TCTACCAT AACTAAAGAT TTAATCTATA TAGCACAATA
19101 TACAATGTTG AGTAATCAT TTTAATTTAT TTTTAACTG ACAAAAATTG
19151 TGCAATATCA TGTATATAT ATATGTATGT GTGTATATAT ATATGATGA
19201 CAACATGATA TTTGATATA TGTATACACT GTGGATGAC TAAATCTATC
19251 AATGGACATG TTTCAATCA CATACTTATC ATTTTGTGT GGTAAAGGCA
19301 TTTAAATCT ACCCTCTAG CAATTTTCAA GTATACAAAT TGTAGTAAC
19351 TCCAATCACA TATTGTACAA TGCACTCTCT AAACATATGC CTCCTGTCTG
19401 ACTGAAATTT TGTATCTTT GACTAAGATC CCGTAATCC CCACTCTCC
19451 CACAGCCCTT GGTAAACACT GTTCTACTCT CTGCTCTCT GAGTTTAAATG
19501 TTTTAGATTT CCACATGTGA GATCATGTG AATTTGTCT TCTGTGCTG
19551 GCTTATTTCA CTTAGCATAA TGTATCCAA ATTCATCTCT GTTGTATAA
19601 ATGACAAAT ATTTGTCTT TCTATGGCTA ATTTGTAGT CATTTGTTAT
19651 ATATATACCA TGTTCCTTT ATCCATTTAT CCAATGATG ACACCTAATG
19701 TGATTTCTAT ATCTGGGCTA TTTGGAATA TCTGCAATG AACATGGGA
19751 TGTAGATGTC TCTTCAATGC ACTGATTTCA TTTGTTTGG TGTATATCC

FIGURE 3, page 6 of 19

WO 02/34922

11/23

PCT/US01/42528

```

19801 AGAAGTGGAA TTGCTGCAIC ATATGGTAGT TCTATTTTAA ATTTTTCGAG
19851 GAAAGTCCGT ACAATTTTCC ATATGGCTGT ACTAATTTAC ATTCCAACCA
19901 AAAGTGTAA AAGGTTCTGT TTCTCCACA TCCTCACCAA CATTTGTCTT
19951 TTGGTAATA ADCATTCTAA TGAGCATGAG GTGATGTCTC ATTATGGTTT
20001 TAATTTAAGT TTCCCTGATG ATTAGTGATG TGGAGCAITG TTTTAAATAC
20051 CTGCTGGCCA TTCAATGCTT CTATGTAGGA ATGTTATTTT AGGTTTTCCT
20101 CATTTTAAAT TCTAGTTATT TGTTTCTTTG CTTTGAATTT GTGTAGTTTC
20151 CTATATATT TTGAATATTA AGCCCTTATC AGATGTATCA TTGCGAGACA
20201 TGTCTCCCA TCCTTTAAGT TGTCTCTTCA CTATGTTGAT TGTTTCTTTT
20251 GTTGTGCAAG AGCTTTTATG TTGCTGCAAA AACCATTTAT CTATTTTTC
20301 TTCTGTTGAC TATACTTCCA GAGTTGTATC CAAAATATCA TTGCCAAGAA
20351 TAATATCAAG AAGCTTTTCT CTATGTTTTT TCTAGTAGT TTTATAGTTT
20401 CAGGTATAT GTTTAAATCT TTAATCCATT TTATGTTGAT TTTGTATAT
20451 GAGGTGAGAT AAAGGTCCAC TTTTATCTT CTACTAGTGC ATATCCAGTT
20501 TTCTCAACAC CATTTATTTA AGATACTGCC CTTTACCAC TGTATGTTAC
20551 TGGAACTTT GTAGATCAGT TGACAATAAA TGTGTGGGTG TATTTCTGGA
20601 CTCTTTATCC TGTTTTATTA GTTTAAATGT CTCTTTTTT AGAAGCTCTA
20651 TGTGTTTTTG GTGACTAGAG CTCTAGTATC AATTTGAGAT CAGGTAGTAT
20701 GATGCACTCC AGCTTTGCTC TTTTGTGCA AAATGCTTT GGCTATTTGA
20751 GTTTTTTAT TCCATAGCAA TTTTAGGGCT TTTTTTTT TCGATTACT
20801 GTGAATAATG CCAITGGAA TTTGATGGAG ATTGCATTGA ATCTTGGGT
20851 AGTATGGAAT TTTTACAGT ATTAATGCT CCAATTAATG AACACAGGT
20901 ATTTTGAAT TTGTGTTTT TCAATTTCT TTCAACAGT TTTTCTT
20951 AATTTAATG TTTTATTTCC ATAGGTTTT GGTAAACAGT GGTGTTTGT
21001 TATGAGTAAG TTCTTTAGTG GTGATTTGT AGATTTGAT GCCACCATCA
21051 CCTAAGCAGT ATACACTGTA CCAATTTGT AGCTTTGAT CCCTCACCTC
21101 CCTCCACACA TTCCCCCAA GTCCCAAG TCCATTTGAT CATCTTATG
21151 CCTTGCATC CTCTAGCTT AGCTCCCACT TATGAGTGAG AACATATAAT
21201 GTTGGTCTT CCAITTTCTA GTTACTCAT TTAGAATAT GGTCTCAAT
21251 TCCATCCAGA TTGCTGGGAA TGCTTTAT TTTTCTCTT TCATGGCTGA
21301 GTAGTATTC ATAGTATATA CATCCACAA TTTCTTTATC CATCTTGAT
21351 TGATGGGCA TTTGACTGTT TCCATGCTT TACAATTTGG AATTTGCTG
21401 CTACAAACAT GCAGGTGCAA GTGCTTTT CATATAATGA CTCTCTTCC
21451 TCTGGTAGA TACCTGTAG TGGGATTTG GGTCAATATG GTAGTTCTAC
21501 TTTAGTTCT TTAAGGAAT TCRCACGT TTCCATAGT GGTGTTACTA
21551 GTTACATTC CCACCAACAG TGTAGAGTG TTCCCTGTT ACTGTATCCA
21601 CACCATCATC TATATTTAT TGATTTTTG ATATGGCCA TTCTTGACAG
21651 AGTAAGGTGG TATTGCACTG TGGTTTTGAT TTGCAITTT CCATCATTA
21701 GTGATGTTGA CCAITTTTTC ATATATTTGT TGGCCATTTG TACATCTCT
21751 TTTGAGAAAT GTCTATTCAT GTCTTTGTC CATTTTGA TGGGATTTT
21801 TGTTTTTTC TTCTAATTT GAGTTCCTG TAGATTTGG AJATTAGACC
21851 TTTGTTGGAT GTGTAGGTTG TGARGATTT CTCCACTCT TTGGGTGTC
21901 TGTTTACTCT GCTGATTTAT TCTTTGCTG TGCAGAACT TTTTAGTTTA
21951 ATTAAGTCCC ACCTATTTAT CTTTGTGTG TGTGTTTTT TTGGGTTGT
22001 TTTGTTTTGG CTGTTGTTG CATCTGCTT TGGGTCTTG GTCATGAAGT
22051 CTTTGCTTAA GCAATATCT AGAAGGTTT TTCTGATGT CTAGAATTTT
22101 TATGTTTCA GTCTTAGAT TAAGTCTTG ATCCATCTG AGTTGATTTT
22151 TGTATAAGGT GAGAGTGA GATCCACTT CATGCTCTA CATGTGCTT
22201 OCCAATATC CCAATACAT TTGTTGAATA GGGTTAATAT TTAAGCTTT
22251 ATATATTTAG GTGTTCTAT TTTGGGTACA TATTTATTTA CAATATCAT
22301 ATCTCTCTGA TGGATTGACC CCTTTCTCAT TATATAATGG TCTTCTGTC
22351 TCTTTTACA GTTTTGTCT TAAAGCCTAA TTTGCTGAT AAAAGTTGAG
22401 CTACCTTTGC TCTTTTTTG TTTCTATTG CATGGAATAT TTTTTCCAA
22451 CCTTCCCAT TCACTCTATG TGTGTTCTA AAGATGAAT GAGATGCTGT
22501 AGGGGCATAT GCTTGGGCT TGTTTTATC ATTCATTAG CCACTCTTT
22551 GATTAGAGAA TTTAATTTAT TTGATTTCAA GGTAAATTT GACAGACAAG
22601 GACTTACTAC TGCCATTTG TTAATTTGTT TCTTGAATTT TTATAGATCT
22651 TTTGTTCTT TCACTCTCT TTAATTTGTT TCTTGAATTT TTATAGATCT
22701 TCTCTAGTGG TGTACTTTGA TTTTACTTT TTAATTTGTT TGTCTACT
22751 ATAGGTTTT GTTTGTTGT TAOCATGAG GTTACATAAA GCATAGTTAT
22801 AAAGGCTAT TTTAACTGA TAACAGCTTA ACTTTCAACA CTTAAATAA
22851 CTATACACTT TTAATCTACC AACTGCCCTC CATTTTATGT CTTTGAATG
22901 ATATTTTACC TATTTTGGG GATGTTGCT CTTATTTGTT ATCCCTTAA
22951 AAATTTATGT AGCAACAGTC ATTTTAAATA GTTTTGGCTT TTAATTTAT
23001 ACTAGAGATA GAATTAATTA ACATACCACC ACTACATTAT TAGGGTATTC
23051 TAAATTGACT ATGTATTTAC CTTTACAGT GAGATTTTT TTTCAATTT

```

FIGURE 3, page 7 of 19

WO 02/34922

12/23

PCT/US01/42528

23101 TCATGTGTGTT AATTAGTATF CTTTCATTTT AACTTGGAGA ATTACATTA
23151 GCATTTTITG TAAGATGGGT CTAGTAGTGG TGAACACCCY CAACTTTITG
23201 TTATCTGGAG ATGTCTTTAC CTCTGCTTCA TTTTGAATA TAACTTTITG
23251 TCATGATTTG AATTGGACAA AATTGTTTTT TTAATTATGC AAAGTGGCAG
23301 GGTAGCAGA ATTACTCTTT TTTTTTTTTT CTGAGACCGA GTTTCACCTC
23351 TCTTGGCCAG CTTGGAGTGC AGTGGCCGAA YCTCTCAGCT TACCGCAACC
23401 TCTGCTCTCC AGGTTCAGGC GATTCTTCTG CCTCAGCCTT OCTGAGTACC
23451 TGGGATTACA GGCATGCAAC ACCATGCTCG GCTAATTTTG CATTTTITAGT
23501 AGAGACGGGG TTTCTCCATG TTGCTCAGGC TGGCTTTGAA CACCGACCT
23551 CAGATGATCC GCGGACCTAG GCCTCCCAAA GTGCTGGGAT TGCAGGTGTG
23601 AGCCACTGCG CCTGGCCAGA ATTACTCTTA TTTATCCTGA GCTTGAGGAA
23651 GAAAGAAATC AAAATTAAAA TTTACATTA CCTAATGGCC AAAGCCTGCA
23701 TTTCAAAATA GTAAATCAGAA AAACATATA AAACACATA AGATAAACAG
23751 ACTAAATATA TGCAATCATT TTAATGAAAC AATCTGACTA GATTGGATGC
23801 AGACTAGGTA GGATGCAAAAT TTAATAAATA CTTTATCTT CTTCACCTTA
23851 TAAACTTTAA AACTGCTTTG TGGAGCAAGT TCTTTTATC TCTGGGGAAA
23901 GATCCTGAGT AAGTCTCATA GAGTCTCAT TCAATTAAT CACAAGAAAC
23951 ATCTTAGTTC AGTAATTAAT CTATCTGGCC CAGTGAATA CTGAAGCTTT
24001 CAATACTTA TCCACTTGAG CTCTCTTTC CATCCAGCT TGGTACTTCT
24051 TTGCTCTAG AAGCCAGCAG TGGTTTATCA TCGACTTAT CTACTGACT
24101 AGCTCCCAA TACCCAGTAG CTGCTGTTTC TGGCCCTCC AGGAATGTT
24151 TTAGGAGGAA AGGGGATAG GAGTAAAGGG CTGGTACTAT TGTATCATG
24201 CCAAGGGCT TGGTGGATAT TCCATGCTTC CCTTCTCTC AAGAGGAAAC
24251 TCCCTTCTT GAGAGCTCTC TCACATAGAC TTTCCAGAGG TCAATCAGGG
24301 GACAAAGAAA TAATGTCTT TAGGCAGACT CTTTTCAGG CTGGTCCCG
24351 AGCTTTCCTT CTGCTGCTGT AATGGTTTA AGGACACAGT TGCATCCT
24401 TGGCTTGGCT CTGCTGCTGT CCTGCTGCTT TCTGTCTGT CTGAGTATA
24451 GCTTTTACA TCAGTCTGT ACTCCCAAA CTCCAAGGAG CACAGTCTAG
24501 ATCATCTAAG TGATCTCTT GAAGCCTCTT GTTAAAGTG GGGGAAGCAC
24551 CCTTCTTCT CCATGGCACT CTGCTTCTC AACCAACTT TAAATAATTT
24601 TTTCTCTCAA AATTCTTANG CTCTCTCTCT TTAATCTCTT GCCATTTTAA
24651 TGTATTATTA CTTTATATGA TGAGCTAAGA GTTACAAAC TGGTPTTAG
24701 AAATCTCTCT AGCAATGTT TTAATGCTAG TTTAGCAGCT CACTTTATTA
24751 TAAGGATATA TGATATATT CTTTGGTTC TCTGCTCTG GGACTCTAGC
24801 TCATCTCTAG GCAGAGAGTC CCATTTTAACT ATTCTGTAC ATAAACAGT
24851 GGCATAATGG CTTTAACTTG AGGGTAATTA TTACAGGAA CAAACAGAA
24901 ACAGAAAAAA AGTAAACTGG TTAATGATATC TGAGTCCCTT CCTTCTCTCA
24951 TCTTACAGG GACCACTGG GTGAGATGTC GTACACCACA ATGTGCATCA
25001 AGGAGAGCTG CCGATTGATT CCTGCACTCC CGTCCATTT CAGAGATCTC
25051 AGCAAGCCAC TTACCTTCCC AGATGGATGC ACATTTGCTG CAGGTCTTTA
25101 CATCTTTTC CTAAGCAGTT CTTAGAGGCT ATGGGATCTT GGAGACCAAC
25151 GTGACAAAGA TTAGTGAGTC TCTTAGACTT TGGAGAGTC AAAAGATAAT
25201 GCTAACATGT GACTTAGGTT TTAATCCTTA TGAGGAGTCT AGAGGATAAT
25251 GCTTTGGTCA GACATGAATT TCAATGACTT TCCCAAGGC ACATAGCCAG
25301 TTGCAGCAAA GCTAAGCCCA GAATCCATGT CTCTGGAATC CCAGGCCAGG
25351 GTCTCTTCCA TTGTGGGACA TCAATTTTAA GATAATCTTT GTTTGGCTGA
25401 GTTGTAGACC GAGCTGAAT TTAATGGAAA ATAGCACCCAG CATCTTTATC
25451 TGAAGACCA AGGGGGATCT TTGCTCTCAT CATCATATA TCACCTTAT
25501 AAATATACAA CATTTAATAG TTAATATAGA GCCTTCAGAC CCATTTATC
25551 ATTTTCCCC TTGGAATCCA ATGTTAACAG ATGCTTATAC AATGATTTAC
25601 AGTTACTGA ACCTTTTAA GTACTTTCAA TGTGGCCAA AATCCAGAGG
25651 CAGCCCAAT GTGTAGATGA CATTAATGA TGTGAGCAGA GCTAGAACTT
25701 GTGCGGAGAC CTTGAGTCTG GAGCCTAGAG TTCTTCCGAA CAACACAGGT
25751 TTTGAGCAG GCTTATAGG AAGCAGAGGG GTCATGTGAG ACATATTATC
25801 TGAATCAATG TTTTATTAA TCAATGCTTA GGAAGCAAGC CAACAGGATT
25851 GCTTCTGGCA AACACCTACA GCTGTACT GTAACTTTGC TGACAGACCC
25901 AGAATTAAT TCTGGAAGCT AGAATTAAT CTGGAACCA AATAACCTC
25951 ACATTTCTCT TCTTTGTTT TGTACTCTGT TTTCCCAAA ACCACATGGA
26001 TATTTGCCAA AATTCTCCAC TTTCCATATG TGAATAGCAC CAATGGAAAT
26051 TTTCTATGGG ATCTGCATGA CAGATCACA GTTCTGTCTG TGTGTGTGTG
26101 CGTCTTCTCT TCAAGACAGA GTCTGTCTAT GTAGCCAGG CTGGAGTACA
26151 GTGCGTAAT CTGCGCTCAC TGCACCTCT GCCTCCAGG TTTAAGCAGT
26201 TCTCTGCTCT CAGCCTCCCG AGTAGCTGGC ATTACAGGTG CACACACCC
26251 CTGCGAAAT TTTTATTTT TATAGAGAT GGGGTTTAC CATGTTGCC
26301 AGGCTAGTCT CAAGCTCTCT ATCTGAGAC CAGCCTCTCT CAGCCTCCTA
26351 AAGCGCTGGG ACTACAGCCA TGAGCCACTG CAGCCAGCCA GTTCTGTGCT

FIGURE 3, page 8 of 19

WO 02/34922

13/23

PCT/US01/42528

26401 TTATACCTA AATTGTCTCC AGGAGTGCTT AATAGTCCAT TAATAGSTAT
26451 TTAGGCCAGG CACAGTGGCT GAOGCATATA ATCCCAATAT TTTGTGACAC
26501 CAAGGTGGGA AGACTGCTTG AAGTTAGGAG TCTGAGACTA GCCTGGGCCAA
26551 CATAGGGAGA CCTGTCTTT ACNAAAAAAA NAAGAGAGA GATAGCCAGG
26601 CATGGTGTG CATGCTGTGA TTCTGCTTA CTGGGGGAC TGAGGCAGGA
26651 GGATCACTTG AGCTCAGAAG TTCAAGGTTA CCCTGAGCAA TGTTCAGGCC
26701 ACTGCTCTCC AGCTGATTG ACAGGCCAGA CCCTGACTCT AAACAAAAAC
26751 AAAAACAATA TATTTAAGTA ATTTCCAAAC ATAGCAGAAA ATATAAGCAT
26801 GGTATTATCAC TTGTATATGA CAOCAACAGC TACTTAAGAT AGACTCATGA
26851 ATTCACTAAA TTGTGTGTG GAAAGCTAAG GTGCCAACCC AAGCCGCTC
26901 TTCTTAGGTG CTCTCACTG GTGTCACTAG CTACAGCAGG CAGAGCATTG
26951 CCAGGAGCTA GCTCTTCCCT TCAAGAACA AAGTCTTGT TAAAGACACA
27001 GTAGCCACA ACTTGCTCTT TCTCTGCAG TCTCTTTAT TTCCCTCCTT
27051 TCTTAGGGAT CACCGTGGT CTTAGTATT GGGGTCTTCA CCACAACCTT
27101 GCTGTCTGGA AAAACCCAAA GGTATGATT TCTCTGTAC ATAAATACTT
27151 CCAAGAACTA ATGCTGTGCA AGTCACATT TGGTAGCTAA GCACAGAAGT
27201 GGCTATATAA TTAAGGGAAA TGACACAAAT TAAACAAAA TAAACATAAA
27251 AGCCAAAAGA AATGTAAAA TATTCTATG TCTTGAACA CTCTTGACGT
27301 GTATCACTGA TTCTTTTCT GTAGGCCACT AAGGTTAAG ATCTATTACT
27351 TGTAAACAGA AGCTGGAGTA TATGTCTCTG TAATAATTGG CCACATCATC
27401 ATTTTGACTT GATTCTAAG TGGATGCACA TCCATTCTA AGTGGATGTA
27451 TCTCCATAGT GAAATAATA CCACTTGCCA TAGTATTTT GTTGGCTGG
27501 GTATCAGACA AATCAGCTGT GAAGCTGCAA GGTCTGCAGG TCTGAAGTA
27551 CACTGCCAGG TGTAGTAGCC ACGGCCACA TADGGCTACT GAGCAGATGA
27601 CATGTGGCCA GTTGAATTG AGTTGTGCTG TANGTTTAA ATACGTGCTG
27651 GATTTTGAAG ACATAGTACC CTAAAAAAT GTGAAACATT TCCTTTTATG
27701 AATTATTTAT ATTGATTACA GGTTCGAATG GTAAATTTG GTTAAATAAA
27751 CTCTATTAG ATTAACCTCA CCTTTTAAA ATGTGAOCAC CAGAACATT
27801 TAAATTACAC ATGTAGATCA CATTATATT CTATTGATCG GTCTAGGTG
27851 GTAGGTGAAG AATGTGTTT ATGTGTTTG GGGATGGTG TGGGGTTGT
27901 CCTCTCATTT CAGGTCTTT ACCCTTGAG GTTCTCTAG GAGAAFTCTG
27951 ATCAGAGACA CCGCTATGCC TACTTACCAT TCTCAGCTGG ATCAAGGTGA
28001 GAACAATTG AAGTTGCTGA AAGTACCCAA AGATGTTTAC TCGAGATAG
28051 TTATTTCTTT TCAGCTCTCT AGCTCTATAC ATTCTTCCAG GGAACCGTAG
28101 ATCTTGCTGC CTATTTGAGC CCCAAGGAT CAGTTAGTTT TACAAAGGAC
28151 AATCGTATT TCTGTACAT CCTTTTGGC CATGCCCTCA AAGCAGTCCC
28201 ACAATGTAA CTACTGCTCA TAGGCTCAAT GCAGTCCACC TCAAGGCAA
28251 GAGAAATAAT TTCTAGATG ACTCCAACTG CCGCTTGTG ATAGGGAAGG
28301 CATCATGTTG GAGCCTCCCA GCTCAAAATC TCACAGTGAA CAATTTAAGT
28351 CTAAAGTTCA AAAGTTTCAA TGGCATTGG TGGAAAAAAT ATCACTTTAC
28401 TGTGACTTC AGACTTCTG TACTAGTAT TACTATAGT CAGAAAGAAC
28451 ATCAITTTTT CAAGTATCAC TTCTTTTCC TCTGTCTTC AGGAACTGCA
28501 TTGGGCAGGA GTTGGCATG ATTGAGTTAA AGGTAACCAT TGCCCTGATT
28551 CTGCTOCAC TCCAGGTGAC TCCAGACCCC ACCAGGCCCT TACTTTTCCC
28601 CACCATTTT ATCTCAAGC CCAAGATGG GATGTATTG CACCTGAAGA
28651 AACTCTCTGA ATGTTAGATC TCAGGTATCA ATGATTAAAC GTACTTTCTT
28701 TTTGGAAGTT AATTTACAG CTAAATGATCC AAGCAGTAG AAAGGGATCA
28751 ATGTATGGTG GGAAGATTGG AGGTTGGTGG GATAGGGCTC TCTGTGAAGA
28801 GATCCAAAAT CATTCTAGG TACACAGTGT GTCAAGTAGA TCTGTTCTA
28851 TATAACTTTG GGAGATTTT AGATCTTTT TGTAAACTT TCACTACTAT
28901 TAATGCTGTA TACACCAATA GACTTTCTA TATTTCTGT TGTTTTAAA
28951 ATAGTTTCA GAATTTATGCA AGTAATAGT GCATGTATGC TCACTGTCAA
29001 AATTTCCCAA CACTAGAAAA TCAATGAGAA TAAAAATTT AATCTCACT
29051 TCACTTAGCC GACATTTCCAT GCCCTGACCA ATCTCTACTG TTTTCTTAAA
29101 AACAGAAATA TTTGGTGTGC ATTCTTTCTG ACTTTTCTT ATACATTTTA
29151 TATGTAGAAA TGTAGCAATG TATTTGTATA GATGTGATCA TTCTATATT
29201 GTTATTGATT TTTTCACTT AATAAAAAAT CACCTTATTC CTTATCATIG
29251 CTATTAGGTA TTCTGTAAAT TGAATGACT ATAAATTTAT TAACTATTTT
29301 CCTATTGGG CATTAAAGTT ATTTCTAGTT TTAATAACAT GCTTGTCAAT
29351 GGCACAAAA GCAAAAAATG ACAAATGGGA TCTAATTTAA CTAAAGAGCT
29401 TCTGCACAGC AAAACAAACT ACCATCACAC TCAATGGGCA GCCTACAGAA
29451 TGGCAGAAAA TTTTGCACAC CTACTCATCT GACAAAGGCC TAAATATCCAG
29501 AATCTACAT GAACCTCAAC AATGTACAA GAAAAAACA ACCCATCAA
29551 AAGTGGGTG AAGGATATGA ACAGACACTT CTCAAAAGAA GACATTTACG
29601 CAGCCAAAAG ACACATGAAA AATGCTTAT CGTCACTGGC CATCAGAGAA
29651 ATGCAATCA AAACCAAAAT GAGATACCAT CTCACACCAG TTAGAATGGC

FIGURE 3, page 9 of 19

WO 02/4922 14/23 PCT/US01/42528

29701 AATCATTAAA AAGTCAGGAA ACAACAGGTG CTGGAGAGGA TGTGGAGAAA
29751 TAGGAAGACT TTTACACTGT TGGTGGCAGG AGAATCACTT GAACCCGGGA
29801 GGGGGAGGTT GCAGTGAGCC GAGGTGGCGC CACTGCACTC CAGCCTGGGC
29831 GACAGAGCGA GTACTCCATC TCAAAAAAAA AAAAAAAGGA CACCAACTTT
29901 CTCATCTTAA ATGTTGTCAT CTATGTGTTA TCTTCCATAA TCTCTCTCAG
29951 ACAGAGTCAT CTTTGTCTGA TATGATCTTA CAGTATTTTT TGTATTATAC
30001 ATTATAATCT CATTAAATTG AGCAACACAA ATGACAAAAG ACAACTGATT
30051 TCTCCCTTTG GATGACCTAA TTTGCTTTCA CTCTTCCATC ATCACTTATA
30101 ACATGATGAT TCTCAAAATC ATCTAOCCTA AATCTATATA TAAAAAAATC
30151 CTTCCCTTGA ATTCCAGATC CTTGGAGACA AACACCCAGC TCTAAAACCA
30201 AATTTGTTTA ACACGGGACC AGTGGTCCCTG TGTGACTTTC CATTTTGTCA
30251 CTATTTTGTG AGCTGGTATA CCAATATCCA CCGAGTTAAA CAATATTTCC
30301 TTGTTTITTT CTGGTACAAA CCAAAATAAA TTACAAACAT CAATAAAGAT
30351 AAAATTTCTA AATAACTCAC TTTCTCTATA TATCTCTTTC TTGCTGGAAA
30401 AATGGGTTAG GTTAGTCTTT TAAAAGCATG CATGATAAAT TGTACTGAAT
30451 ACAATTTTCA GGTCTGGACA TACTAGGTAT AATTTTCTGT GTCTCTGGGG
30501 TCTTACCTAT TTGGGGTCAA AATAAACAGG TTTATTAAAG TTATTAAATAT
30551 TCAATTTTCA TATCTTCTTT AACAAATTATG TTTCTGTTGA GTTTCATTGC
30601 CAATAATTTA TTTGTCAGGT TGCCAGGTCG TTTTAAACTT CTGTGTATTT
30651 TTTCAATATC AATTTTACTT TAAATATTTT TACAAAAGAG GTCTGTAAAA
30701 TTTCTAATA ATTATTATAT TATTGTTTTT TCACTGACAT TTTGTGAATT
30751 GAAAACCCCT AAAAATATGA AATCATTTTT TCGAAATATG TGCCACAGAC
30801 AATTTTGTGA AATAAGAAGA CAGAAACAGG GCATTATCAA GAGATAAATA
30851 TTCAATATAC CTTATATTTT TGTACACAT TTTTATACCA ACTGTGCCAA
30901 AAATTTGATA TCATATAAAT GATAACAGT TCACAAAGGC ATTCTTTTAT
30951 CCTTAACTC TCAAAATTGA AACTTTCTATA GGTAGGAAGT AGGGGAAGCA
31001 TATATTCCCT TTGAAAGGTG CAAGAAATATG TCATTGGCAT TCACCATGGT
31051 ACTCTTCAGG CTTAAAAAAA ATGACTGTCA AAACATTATC AAACATAGCA
31101 TATTTATTGG GTACCTTTAT GTTTACATAA ATATTGAAGA TATCTCACAT
31151 ACCTCTTCA ATCAGATTAT CTCACTGACA TTTATTGAOC ACTTTCTATG
31201 GGGAAAC

FEATURES:
Start: 2191
Exon: 2191-2367
Intron: 2368-8318
Exon: 8319-8460
Intron: 8461-9761
Exon: 9762-9806
Intron: 9807-11566
Exon: 11567-11694
Intron: 11695-14298
Exon: 14299-14426
Intron: 14427-14509
Exon: 14510-14664
Intron: 14665-17152
Exon: 17153-17259
Intron: 17260-17834
Exon: 17835-18025
Intron: 18026-24959
Exon: 24960-25093
Intron: 25094-27056
Exon: 27057-27121
Intron: 27122-27913
Exon: 27914-27996
Intron: 27997-28492
Exon: 28493-28664
Stop: 28665

CHROMOSOME MAP POSITION:
Chromosome 1

ALLELIC VARIANTS (SNPs):
DNA
Position Major Minor

FIGURE 3, page 10 of 19

WO 02/34922

15/23

PCT/US01/42528

| | | |
|-------|---|-----|
| 267 | T | C |
| 284 | G | A |
| 1269 | T | C |
| 2487 | T | C G |
| 4186 | G | A |
| 4522 | G | A |
| 4522 | C | A |
| 5075 | T | G C |
| 5450 | T | C |
| 5450 | T | C |
| 5995 | G | A |
| 6241 | G | A |
| 8479 | C | T |
| 10045 | C | A |
| 10045 | G | A |
| 11994 | G | A |
| 14070 | A | G T |
| 15535 | T | C |
| 17618 | C | T A |
| 18520 | A | - C |
| 18525 | - | T A |
| 18525 | - | G A |
| 19189 | T | C A |
| 19259 | C | T |
| 19325 | G | T |
| 19346 | G | T |
| 20843 | - | T |
| 20845 | T | C |
| 22234 | T | C |
| 22234 | G | T |
| 22247 | C | T |
| 22334 | A | G |
| 23033 | T | - |
| 23036 | - | A |
| 23421 | A | C |
| 25582 | T | C |
| 26407 | C | A |
| 26473 | C | T |
| 26844 | G | A |
| 28384 | A | - |
| 28417 | A | C |
| 29265 | A | O |
| 29484 | A | G |
| 30417 | T | - |
| 30783 | C | G |

Context:

DNA

Position

267

CCAGGCTCTCTTAGGCTCCTAAATATAGTGCAAAAAGTTCCAGAGTTCCTTTGTTACCCA
TGAAAGCAGATGGAACGGTGTCTGGACAGGGGCAACTGGCCCTGGAGCAGAGGAGTAACCTG
CATAGAACTGTCCAAGCCTCAGAGGGAGTCAACACCACCAAGCAAGAACCTGGGTGGGAGTA
GGTGAGCCAAGGGGTTCCAGGCTCTGACCTGCCAAGAGAACTCATTAGAAGGTCAACA
ACCACACATCTATTCTCGGTCTCA

[T, C]

GAAGAAGCCAGGGACCGGACAGGCAAGATATCAAAAGCTGAAGTTTCAGCTCTGGGC
AGAGCATGGATCTGAGGCTCTTTGGCCCTACCAOCATGGCATCATATGAGGGCCATCATAC
AACCATCATGATTTGGGGGAGGAATAGGGCATAGAGGAATCATATGAAAAGCTGAAATGC
CATGAGTTACCCAGAAAGCTGTGTAAAGCCAGAGGATTCGAGACCTGTCAATTAACA
ACATCTAGTTGAAGGTTGGAGTTAGGTAGGAGGTAGGGAAGTCTGGCAAGAAAGGAGCTG

284

CCAGGCTCTCTTAGGCTCCTAAATATAGTGCAAAAAGTTCCAGAGTTCCTTTGTTACCCA
TGAAAGCAGATGGAACGGTGTCTGGACAGGGGCAACTGGCCCTGGAGCAGAGGAGTAACCTG
CATAGAACTGTCCAAGCCTCAGAGGGAGTCAACACCACCAAGCAAGAACCTGGGTGGGAGTA
GGTGAGCCAAGGGGTTCCAGGCTCTGACCTGCCAAGAGAACTCATTAGAAGGTCAACA

FIGURE 3, page 11 of 19

WO 02/34922

16/23

PCT/US01/42528

ACCACACATACTATTCCTGGTCTCATGAAGAACCCAGGGACC
 [G, A]
 GACCCAGGCAAGATATCACAAAGCTGAAGTTTCAGCTCTGGGGCAGAGCATGGATCTGAGG
 TCTTTGGCCCTACCCATGCGATCATATGAGGGCCATACAAACCATCATGATTGGG
 GGAGGAATAGGGCATAGAGGAATCATATGAAAGCTGAATGCCATGAGTTAOCOCAGAG
 AACCTGTGTAAAGCCAGAGGATTCTGAGACCCCTGTCAATTAACAACTCTAGTTGAAGGTT
 GGAGTTAGGTAGGAGGTAGGGAAGTCTGGGAAGAAGGAGCTGAACACTTGTGTGTGT
 1269 CCTGTCTTAATCACTTACCCGCCAAATAAATCTGGCTCCAGAGAGTGGAGCGTAGGCTT
 AAGCAATTGGGGCCGGAAGGGCCGGAAGCTGGGGCAGGCACAGTATAGGGAGAACAGG
 GAATTGTAGCAGAAATTGGGTTATTGTTTCAGAGCTGTCAATGAACACTTAACATATGCC
 TGTCTTAGCCTAAATCAATGAATAAATGAATGAATAAATGAATGAATAAATGAATGAATAA
 TGCCATATAAGATGTCTGGGACAGGAGGTGGGGGAGACACCAGCTTGGGAAGTCAGG
 [T, C]
 TGTAGATCCTAGTTACCCACCTGATACGTTACAAATACTAAACCATCACTTTCAAT
 ATTTTACTACATTTCTGTATCTGTACTCGAGTTTATTTATGTTTCTGGCATCTAGA
 GTGAGCCCTTCATGGGCATGAGACCCAGCCAGCCAGAGGCTCTGAACCCAGAGAGC
 ATATGCTCGGTTTAAATGCTGTCTCATTTAGAAATGTTAATAAAGTTTATATCCGCAT
 TTTCAATTTGCACTGAGATTCAATAATATATAGCAGGCGCTGACTGTACTGTATAGTGG
 2487 AGCATTGGAAATTCCTCTGGCTGGAGAGCGGCTGGGGGGGGGCTTTTACCTGGCCTTCGT
 GTTCTGCTGGCCCTGGGGCTGCTGAGGGCAATTAAGCTGTACTGGGAGGAGCGGCT
 GCTGGGGAGCTGGGGCCCTTCCAGCGCCCTCCACCCATGCTTCTTGGGACCCAGAA
 GGTAAATGGAAGGGAAGAGGTTAGAAAGGAGGAGAGGCGGCGGAGGAGGATGCGGC
 AGAGGAGCGGAGCGGAGAGAGAGCGAGCTTTCTTCCATCCCTGGGAGCCTTCCGCTT
 [T, C, G]
 CACCGGCTTTCCAGCCCGGCTGTGGCTCTTAGCATCATTTCTCTGCTCTGGAGAAT
 TGCTTTCCCGCAGCCCTCAGGGAAAGGTCACAAAGAGGAACTTTGGGGCTGGGAGA
 GAGCTATTTAAGAACCTGAATATGGAAGAAAGAGGAGCTGTAACTCAAGTCTGTCTC
 TCAATGCTTCAACAGCCTTCCACATGCTGTGCTTTAAATAGCATGTTATCTTAAATA
 ACTTATTAGTTGCAAGAAATATGCAAAATCTATCCCAATCGTTGGCAGCCTTAGTCCAAT
 4486 TGTATGTATCTCTACTGTCTCATGAATACTATGTGCTGTGTGTTTAAATGAATGTT
 TTGGCATCCTGTCAAAATCAATTTGACCAATAATGTCAAGGCTATTTCTGAGTCTTCA
 ATTCTAATCCATTGATCTATATGCTATCTTAACTCATGGACAGAGAGTAGAAGGATG
 GTTACCAAGGCTGGGAAGGATAGAGGGAGCTGGGGGAGGAGTAGGGAAGGTTAATGG
 GTACAAAAAATAGAAAGATGAATAACACCTACTATTTGATAGCATAGCAGGCTGGCT
 [G, A]
 TAACACCTACTATTTGATAGCATAGCAGGCTGGCTATAGTCAATAAATGTAACCTTT
 TAAATAAAGAGTGTAAATAGGATGTGTTGCAACTCAATGGATAAATGCTTGAAGGATGGG
 TACCCATCTTCTCATGATGTGCTATTTACATTTGCAATGCTGTATCAAAACATCTCAT
 TTACTCCATAAATATATACACCTACTATGTATCCCAAGTATTAATAATTAATAAAT
 AAATATATAGCTATCCTTATGCTAGTACCACTGCTTACTGTGCTTTGATGAAGC
 4522 TGTATGTATCTCTACTGTCTCATGAATACTATGTGCTGTGTGTTTAAATGAATGTT
 TTGGCATCCTGTCAAAATCAATTTGACCAATAATGTCAAGGCTATTTCTGAGTCTTCA
 ATTCTAATCCATTGATCTATATGCTATCTTAACTCATGGACAGAGAGTAGAAGGATG
 GTTACCAAGGCTGGGAAGGATAGAGGGAGCTGGGGGAGGAGTAGGGAAGGTTAATGG
 GTACAAAAAATAGAAAGATGAATAACACCTACTATTTGATAGCATAGCAGGCTGGCT
 [G, A]
 TAGTCAATTAATGTAACCTTCACTTTTAAATAAAGAGTGTAAATAGGATGTTTGCACCTCA
 TGGATAAATGCTTGAAGGAGTGGGTACCCCTTCTTCAATGATGTGCTATTTCACTTGC
 ATGCTGTATCAAAACATCTCATTTACTCCATAAATATATACACCTACTATGTATCCAC
 AAGTATTAATAATTAATAAATTAATATAGCTATCCTTATGCTAGTACCACTG
 CTTACTGTGTGCTTGTAGTAGCTTTGAAATCAGGAAGTATGAGTCCCGGCACTTTGG
 4522 TGTATGTATCTCTACTGTCTCATGAATACTATGTGCTGTGTGTTTAAATGAATGTT
 TTGGCATCCTGTCAAAATCAATTTGACCAATAATGTCAAGGCTATTTCTGAGTCTTCA
 ATTCTAATCCATTGATCTATATGCTATCTTAACTCATGGACAGAGAGTAGAAGGATG
 GTTACCAAGGCTGGGAAGGATAGAGGGAGCTGGGGGAGGAGTAGGGAAGGTTAATGG
 GTACAAAAAATAGAAAGATGAATAACACCTACTATTTGATAGCATAGCAGGCTGGCT
 [C, A]
 TAGTCAATTAATGTAACCTTCACTTTTAAATAAAGAGTGTAAATAGGATGTTTGCACCTCA
 TGGATAAATGCTTGAAGGAGTGGGTACCCCTTCTTCAATGATGTGCTATTTCACTTGC
 ATGCTGTATCAAAACATCTCATTTACTCCATAAATATATACACCTACTATGTATCCAC
 AAGTATTAATAATTAATAAATTAATATAGCTATCCTTATGCTAGTACCACTG
 AAGTATTAATAATTAATAAATTAATATAGCTATCCTTATGCTAGTACCACTG

FIGURE 3, page 12 of 19

WO 02/34922

17/23

PCT/US01/42528

CCTTACTGTTGCTTGTAGTAAGCTTTGAAATCAGGAAGTATGAGTCCCGGCACTTTGG

5075 TTTGTAGTAAGCTTTGAAATCAGGAAGTATGAGTCCCGGCACTTTGGTATTTTCCAAGA
TTATTTTGGCTGTTTGGAAATCCTTGATTTCTATACAAATTTTAGACTCAGCCTATCAAT
TCTACAAGGAACAGCTAGGGTTCTGCTTGGGATTGCACTGAATCTGTAGATCAGTTTG
GGCATTTATGCCATCTTAAGAAATATTAGTCTTCTGATCCATGAACACAGAAAGCCTTTC
CCTTTAGTTAGTCAATCTTAATTTTCTTGTGTTTCTTTTGTGTTTGTAGACAGAGT
(T, G, C)
CTGCTCTGTGCCCCAGGCTGGAGTGCAGTGAAGCAATCTGGCTCACTGCAAGCTCCGGC
TCTGGGATTCAAGGATTTCTGCTGCTCAGCCTCCCAAGCAGCTGGGACTACAGGCACAT
GCCACCAACCAACTAATTTTGTATTTTCAGTAGAGAGCGGGTTTCAACATATTGGCCA
GGCTAGTCTCGAAGCTCTGACCTCTGATCCACCGGCTCAGCCTCCCAAGTCTGGGA
TTACAGGGGTGAGCCACTCCCGGCTTCTTTAATTTTAAAGATGTTTGTAT

5450 GATTCTCCTGCTCAGCCTCCCAAGCAGCTGGGACTACAGGCACATGCCACCAACCAAC
TAATTTTGTATTTTCACTAGAGAGCGGGTTTCAACATATTGGCAGGCTAGTCTCGAAC
TCTGACCTCTGATCCACCGGCTCAGCCTCCCAAGTCTGGGATTACAGGGGTGAGC
CAGCAGTCCCGGCTTCTTTAATTTTAAAGATGTTTGTATTTTCAAGATATAC
ATCTGCAATTTCTTTGTAAATTTAATTTGTGTTGTTCTTTAATTTCAATTCAGACTA
(T, C)
TTATTCGATTCATAGTCTTTTACAGTCCACATTCCTCTGACTGTCACTAAGTTTCTT
TTTCTGTTTTTGGAGGTTTCTATCAGAAATTTGCGAGTACAGAGTACGGGACATGTCA
AAGTCTTAATATTACCAACCTCCCAATTTATCAGATCAGGATCCTTTGGTATTCAC
CATGAGGGGAAATCTAGTATCTAAGGCTCAAAAGTGTACTGTTTACATAGGCAATTA
CATTTTATGCTACATAAATACTACATATTATGGAGTACCTGTATATTGTATACGTG

5450 GATTCTCCTGCTCAGCCTCCCAAGCAGCTGGGACTACAGGCACATGCCACCAACCAAC
TAATTTTGTATTTTCACTAGAGAGCGGGTTTCAACATATTGGCAGGCTAGTCTCGAAC
TCTGACCTCTGATCCACCGGCTCAGCCTCCCAAGTCTGGGATTACAGGGGTGAGC
CAGCAGTCCCGGCTTCTTTAATTTTAAAGATGTTTGTATTTTCAAGATATAC
ATCTGCAATTTCTTTGTAAATTTAATTTGTGTTGTTCTTTAATTTCAATTCAGACTA
(T, C)
TTATTCGATTCATAGTCTTTTACAGTCCACATTCCTCTGACTGTCACTAAGTTTCTT
TTTCTGTTTTTGGAGGTTTCTATCAGAAATTTGCGAGTACAGAGTACGGGACATGTCA
AAGTCTTAATATTACCAACCTCCCAATTTATCAGATCAGGATCCTTTGGTATTCAC
CATGAGGGGAAATCTAGTATCTAAGGCTCAAAAGTGTACTGTTTACATAGGCAATTA
CATTTTATGCTACATAAATACTACATATTATGGAGTACCTGTATATTGTATACGTG

3995 TTATTCGATACATAAATACTACATATTATGGAGTACCTGTATTTTGATAAGTGCATA
CAAGTGCAGTGCATAAATCAGGGTGTAGGGTATTCATCACTTCTAACAATTTATTAAT
TATTTGTGTTTGGAACTTTCAAGTCTCTTCAAGTCTCTCAGAAATATTCAATACATTAAT
TGTTAAGTGTCTATTGAACTCTGAACTTATTCTCTATCTAAGAGCAGTAACTTTT
AAGTATAGTACATAAGGTTACAGAGGATAAAGTGTATAGGGAATTCCTACAGAT
(G, A)
AGAAATTCATTCCTTACTCTTAGTAATACAGGCTCTCAACATGCCAAGGATATTCTCC
CTTGGAGCTTTGAACATGCAGCTCTGTGTTATATTGCTCTCCCTGCAAAATTTCTCTAA
AAGAGGCTTGGCCTGACCATTCAGACTAAATAGCACTCTAGTACTCTCTATCTCAAC
CCTATTATTATTATCTTGGCCCTTATCACTCTCTGACACTATAGTATCTCTTGTGCT
TGTTGTTTATTATCCCACTAATACATATAAATCTGTGAGAGGTAGGATCTTGT

6241 AGTCATAAGGTTACAGAGGATAAAGTGTATAGGGAAATTCCTACAGATGAGAAAT
TTCAATCTTACTCTTAGTAATACAGGCTCTCAACATGCCAAGGATATTCTCCCTGG
AGCTTTGAACATGCAGCTCTGTGTTATATTGCTCTCCCTGCAAAATTTCTCTAAAGAG
GCTTGGCCTGACCATTCAGACTAAATAGCACTCTAGTACTCTCTATCTCAACCTAT
TATTATTATCTTGGCCCTTATCACTCTCTGACACTATAGTATCTCTTTGCTGTTTC
(G, A)
TTTATTATCCCACTAATACAAATATAAATCTGTGAGAGGTAGGATCTTTGTTGCCA
CTATAACCTAGTGCATGTTAGCTTCTGGTGCATTAATAGGTGCTCAATTAATCTCTT
TTGAATGCATAAATATATTAGGTGCTGAGAAATTTATTTATCAAGATCAATTTACTG
CATAGAAAGGCCAGGTGTTTGCATTTATTCAATAGCCACATATGGGACCTAGGATG
TACATATGCAAGTGTGTGTGTATGTGTGTGTGATCTGCAATGTTGATGTTGATG

8479 AAAGCATGTTATGTCACCTCCAGAAAGTCTCAGGCTCTCTGCTGCTGACCTTATCA
GGTCTGAACTCAGCTTGTGCTATAGAGGGGACAGTCCAGCTTGGCTGGCTAATTAAC
TTTACTTTTTCAGTGCAGTTTATTAGGATGATACATGGAGAGGCTTGGGAAATTA
TTGAAATACCTCTGCTTCCCTTCTGGATTGGGCTTTTCAAGCATTTTCTGTA

FIGURE 3, page 13 of 19

WO 02/34922

18/23

PCT/US01/42528

TCTATGACCCAGACTATGCAAGACACTTCTGAGCAGAACAGGTAAGAAGAGGGGGAAG
(C, T)
TCTGGGACCTATTCTCTCTAGAGTGAATGCATAAAACCATAGGCAAGATTCCAAAGC
AAAGATTGGTTTGGGGCTTTAGAGACACAGCAGCAAGTATGGGGAGGTGACAGGTTTC
CTACCAATACGAAGGGGATTCCTATCTCTCCCACTGCTTCTCTTCTTCTGAGTATGC
ATGGGCACGTTGAGTGGGTATACTTAAAGCCTAGCTGGCATTAACAGACTTGGCAGGC
AAGGCTTCCCTTGGCTCTGTGGTCTTATGACTTCAGTGTGAGCAACACTTCCCACTCC

10045 TCTGCTTGACTCTGCAGATCCCAAGTCCAGTACCTGCAGAAATCTCACCTCCACTTCT
TGGTATGTATGTGCAATGAGAGGTATAACCACTCTCATTCAAGTCCCTTTCCATAG
TAGAGCATGCCAAAGAACTGAAATCTGAATTCAAAGCACAAGAGTGCAAGGTAGAGC
TATACTCAACGTTATCTAGGGGAAGATTGAGGGGAGCTCTAAGGTCAACACACCA
CTTCCAGAAAGCTTCTTCACTCGTCTCTCTCCCAAGTCTTATTCTCAAGGCAGCAG
(C, A)
TACATGAATCTGTCCCTCTCTCTTTAAACTACAGCCTTGGCCAGGCACAGTGACTCAT
GCATGTAAATCCAGCACTTTGGGAGGCCAAGGTGGGAGGATCACTTGAGGTCAAGATTTC
AAGACCAAGTGGGCCAATGAGTGAATCCATCTCTACTAAAAATACAAAAATTAGCCA
GGCATGGTAGCATGTAGGCTGT

10045 TCTGCTTGACTCTGCAGATCCCAAGTCCAGTACCTGCAGAAATCTCACCTCCACTTCT
TGGTATGTATGTGCAATGAGAGGTATAACCACTCTCATTCAAGTCCCTTTCCATAG
TAGAGCATGCCAAAGAACTGAAATCTGAATTCAAAGCACAAGAGTGCAAGGTAGAGC
TATACTCAACGTTATCTAGGGGAAGATTGAGGGGAGCTCTAAGGTCAACACACCA
CTTCCAGAAAGCTTCTTCACTCGTCTCTCTCCCAAGTCTTATTCTCAAGGCAGCAG
(G, A)
TACATGAATCTGTCCCTCTCTCTTTAAACTACAGCCTTGGCCAGGCACAGTGACTCAT
GCATGTAAATCCAGCACTTTGGGAGGCCAAGGTGGGAGGATCACTTGAGGTCAAGATTTC
AAGACCAAGTGGGCCAATGAGTGAATCCATCTCTACTAAAAATACAAAAATTAGCCA
GGCATGGTAGCATGTAGGCTGT

11994 GGTAAAGTAAAGGGGGAAGTGTCTGTGCAATTCGGAATGCTCCAGCAATGGACAGTAT
TAGGTATGTGTTTGTGGGCAATGAATAAAAAATCAGTTCTAAAAATTTAAACCAATG
TACAGCTACTTATTAACCAATAGGTGTCTGTAAAAATTTTATATGTTCTTTGAGTGATA
ATATTAAATAAGATCTGGTCTCTGTCTTAGATATATTTTGAATTTTATGGCAGCAA
ACCAAGTACCAATGCTGATAGTATAGTATAGTGTGTGTATGTGTCTCATGGAGGGC
(G, A)
GGTCTGTACAAACCTACCCCAAGTCTGAGGAAGTCTGAGAGGCTGAAGAAAAAGGCTGAC
AGTTTCTTAAAGAAACATTCATAGAGGCTTCAACAAAAACCAT

14070 GGTCAAGCTTTGCTGGGGCAGCTCCCTGCAACAGCTCCTCTCCACACTTGTCTGTCTTC
TCCTTTTGAATCCAAAGCTTTTGAATAATGTTCTGAGTTTATTTAAAAATGTGGCTATG
GTGTTGAGAGCAGTGGCAGGTTACCTAGCAAGTTTGAATTTGAAGTTGAGGAGAGCCCT
GGGGTAAACCCCTTGTAAATATGGGTCTTGTGTCAATGATTGCTTTAATGGAATCTGGT
CTGTTTCAAGCAGAGTTATGGTAAATTAATGAAGCCGAGATCTTTAATCAGCCATT
(A, G, T)
ACCATATATGCAATTTCTCCATGCTCTCTCACTCCGCTGGGTGTATTTTCCCTTCC
TCGTGCCCTGTGTAAGCACATGGCTTATTTACTCATGTGATCTTTGGTTCCTGTGGGT
AGGGTTGTCTCCATTAGATCATAAAAACAGGGCCAGGCAGGAGCTTCAATGAAGGCAA
TTTGGTCAATGGTGGTGGTATGATGTGGTCTTGAATCTCTGTGCCAGGATAAGTGGGAG
AAGATTTGAGCACTCAGGACACAAGGTGGAGGTCTATGAGCAGATCAACTCGATGTCT

15535 ACTTACTGCTTTGTTTCACTGTCTATATAATGGAGATTAAAAAGAACTATCTCAZ
ACATTTGTTGTTAAGATCACTGGTTAATATATATAAGCAATTTAGGACAGTSCCTGGCA
CTGAATAGATGTTAAATGTAAAGTATAGTTATGTCAAAATGCTTTGCTTCCAGGAATTT
GCAAGACACACACATATGCACACTTACACATACATATATGATATATGACATGACATAGATA
TTATAAGAGGACACTCAGAGAGCAGGTTATAACAATTTAAGGCATAAATGGGCATTA
(T, C)
AAATAGCAGCAGTTCCCAAGTCTTTCTGCATCATTGCACACAGAAAAATGTTAATGTTT
TTGTGCTTCAATGGAGTAAACAGCAATGGATTGGGGGAGCTATACAGAACTTTGTAA
AAAAATCTTTACTTTTAAATATTAACAATTTATGATGAAGCAAAATGCAAGTGT
TAGGGAAATATTAATGTTAAATTTATTAACAATTTAAACCTTTTCAATTTTCTTT
TTTTTTTTTGGATCGACTCTATCACTCAGGCTGGAGCGCAGTGGTGTGATCTCAG

17618 GGTAAAGTGGGAATGGGATGGGAGACAAAGATAAACCGATTGACTAAATTTAATGTAC
TTTGAATGATGAGCAGCTTCATGCAATTTGAGACAAAGAGAGATTTCTGCACTGTGTC

FIGURE 3, page 14 of 19

WO 02/34922

19/23

PCT/US01/42528

CCTAGAGGAGGCTTAGTAAAGACTAAACGAAGATTGACAAAGATTGAGCATTGTCATA
 TGGATACATGGATTAGGGGATCATGAAAAAATGGTCACATGGATAAACGTAATAATTA
 TCGATGATAAGGTCTGGGAAATCTGGGAGTTTGAAGAGAAATTTCTAGGGCCTGTTGATCG
 [C, T, A]
 GGGCCCTTTGTGCAAGGCCCTGCTTTCTTATCTAACCTTGGTTCCTCTTATGCTTTGGG
 CAGAAATAGGTTTATACCAATATTTGTTGAAGCTGAATTAATAATTAACCCCTATTTAA
 AGCTCTGATTTTCCCTCAAACTATTATTGTTGTTATCTCCAAACATTTATAACTG
 GCATTTTATTAATAATTTGTATTGTACTTTCTAGGATGAAGTGGTAGCAGCTTCTCA
 GATATTGATGTACACTCTGAAGTGAGCATTCTCTGTTGGCAGGACATGACACTTGGCA
 18520 ATTATCCATAAAATGTCGTCTATTGGTTTCTAATCAATGGTGTGTGAATGTCTTATT
 TCTTTATTTACCTTGGCTCTGATGCAATGGAAATGAGGACTTGTCTCCTGGGCTGGCAC
 TTAGAACTTAAACAAATAGGGTCCAAAGTGGAGCTCCTCTCTGAGAGAGCTGAATGATTAG
 CTGCAATTTTAAAGGCTCATTTTAGACATCTCCAGCCGCTTGTACCAATTTTATTCTCT
 CAGGATTTGATTTAGACTTCAGACATAAATTTGATGATATATCTATAGTTAAGTTTAG
 [A, -, C]
 AAATATGGACTGAGGACATTTAAATACTGAGACTTTTCTATGACTACAAATTTATTGTG
 GGGCCTGTCTTGGGTGAGCTAATGGTCTAATACAGGAGACAGGAGACAGCCTCCAAAT
 GCACTGTAGCAATATGAGGCAATGATAGAGATATGTGCTGGCTAACACAAAGACATAGA
 AGACAGGTAACCTACCTGGCATGGGAGCTCAAGGAGACTTCTTGACATTTACGCTGACT
 GCAGGATAGTAGGAGTTAGCCAGGTGGAACTGTCTATCTCTATCTTGTAGACTTTAAG
 18525 TCCATAAAATGTCGTCTATTGGTTTCTAATCAATGGTGTGTGAATGTCTTATTCTTT
 ATTCTACCTTGGCTCTGATGCAATGGAAATGAGGACTTGTCTCCTGGGCTGGCACTTGA
 ACTTAAACAAATAGGGTCCAAAGTGGAGCTCCTCTCTGAGAGAGCTGAATGATTAGCTGCA
 TTATTTAAGGCTCATTTTAGACATCTCCAGCCGCTTGTACCAATTTTATTCTCTCAGGA
 TTGATTTTAGACTTCAGACATAAATTTGATGATATATCTATAGTTAAGTTTAGCAAAAT
 [-, T, A]
 TGGACTGAGGACATTTAAATACTGAGACTTTTCTATGACTACAAATTTATTGTGGGCC
 TGTCTTGGGTGAGCTAATGGTCTAATACAGGAGACAGGAGACAGCCTCCAAATTTGCACT
 GTAGCAATATGAGGCAATGATAGAGATATGTGCTGGCTAACACAAAGACATAGAAGACA
 GGTACCTACCTGGCATGGGAGCTCAAGGAGACTTCTTGACATTTACGCTGACTGCAGG
 ATAAGTAGGAGTTAGCCAGGTGGAACTGTCTATCTCTATCTTGTAGACTTTAAGCATAT
 18525 TCCATAAAATGTCGTCTATTGGTTTCTAATCAATGGTGTGTGAATGTCTTATTCTTT
 ATTCTACCTTGGCTCTGATGCAATGGAAATGAGGACTTGTCTCCTGGGCTGGCACTTGA
 ACTTAAACAAATAGGGTCCAAAGTGGAGCTCCTCTCTGAGAGAGCTGAATGATTAGCTGCA
 TTATTTAAGGCTCATTTTAGACATCTCCAGCCGCTTGTACCAATTTTATTCTCTCAGGA
 TTGATTTTAGACTTCAGACATAAATTTGATGATATATCTATAGTTAAGTTTAGCAAAAT
 [-, G, A]
 TGGACTGAGGACATTTAAATACTGAGACTTTTCTATGACTACAAATTTATTGTGGGCC
 TGTCTTGGGTGAGCTAATGGTCTAATACAGGAGACAGGAGACAGCCTCCAAATTTGCACT
 GTAGCAATATGAGGCAATGATAGAGATATGTGCTGGCTAACACAAAGACATAGAAGACA
 GGTACCTACCTGGCATGGGAGCTCAAGGAGACTTCTTGACATTTACGCTGACTGCAGG
 ATAAGTAGGAGTTAGCCAGGTGGAACTGTCTATCTCTATCTTGTAGACTTTAAGCATAT
 19189 CTGGTCAAGGGACAGAAAGACAGAAATGCTAAGGACAAATTCAGCAGCAGACAGATAAA
 AAACACCATATTTTCATATGCAAAAGTCAACTCAATTTGAACATTTGTAAACCAAAATTTG
 ACATTTAAAGATATATCAGAGATCTCATTTTAAAGGAATAGAAAGCCCTTCTCTACCA
 TAAACTAAAGATTTAATCTATATAGCAGCAAAATACAAATGTTGAGTAATCAATTTTAAATTT
 ATTTTAACTGACAAAAATTTGTCATATACATGTTATATATATATGTATGTGTGTATAT
 [T, C, A]
 TATATGATGTACAAATGATATTTTGTATATATGTATACACTGTGGAATGACTAAATCTAT
 CAATGGACATGTTCAATTAACCTCATCTTATCATTTTGTGGTAAGGACATTTAAATTC
 TACCTCTTAGCAATTTTCAAGTATACAAATTTGTAGTAATCTCAATCACATTTTGTACA
 ATGCACTCTCTAACTTATGCTCTCTGCTGACTGAAATTTTGTATCTTGTACTAACT
 CCGTGAATCCCTCATTTCTCCACAGCCCTGGTAACCACTGTTCTACTCTCTGCTTCTT
 19259 TTTCTATGCAAAAGTCAACTCAATTTGAACATTTTGTAAACCAAAATTTGACATTTAA
 AGTATATCAGAGATCTCATTTTATAAGGAAATAGAAAGCCCTTCTCTACCAATAACTAAAG
 ATTTAATCTATATAGCAGAAATACAAATTTGAGTAATCATTTTAAATTTATTTTAAAC
 TGACAAAAATTTGTGATATACATGTTATATATATATGTATGTATATATATATGATG
 TACAAACATGATTTTGTATATATGTATACACTGTGGAATGACTAAATCTATCAATGGACA
 [C, T]
 GTTCATTAACCTATCTTATCATTTTGTGGTAAGGACATTTAAATCTACCTCTTA
 GCAATTTTCAAGTATACAAATTTGTAGTAACCTCAATCACATATTTGTACAAATGATCTCC

FIGURE 3, page 15 of 19

WO 02/34922

20/23

PCT/US01/42528

TAAACTTATGGCTCTGCTGACTGAAATTTGTATCTTTGACTAAGATCCCTGTAATC
CCCATCTCTCCACAGCCCTGGTAACCACTGTTCTACTCTCTGCTCTTTGAGTTTAAAT
GTTTTAGATTTCCACATGTGAGATCATGTGGAATTTGTCTTTCTGTGCTGGCTTATTTTC

19325 TCAGAGATCTCATTTTATAAGGAAATAGAGCCCTTTCCCTACCATAACTAAAGATTAA
TCTATATAGCAAAAATACAATGTTGAGTAATCATTTTAAATTTTAACTGACAA
AAATTTGGCATATACATGTTATATATATATGTAATGTTGTTATATATATGATGTAACAC
ATGATATTTGATATATGTAATCACTGTGGAAATGACTAAATCTATCAATGGACATGTTCA
TTAACTCATCTTATCATTTTTTGTGGTAAGSACATTTAAATCTACCTCTTAGCAAT
(G,T)
TTCAAGTATACAAATTTGTTAGTAACCTCAATCACATATTGTACAAATGCATCTCTAACT
TATGCTCTCTGCTGACTGAAATTTGTATCTTTGACTAAGATCCCTGTAATCCCTCAT
TCTCCACAGCCCTGGTAACCACTGTTCTACTCTCTGCTCTTTGAGTTTAAATGTTTAA
GATTTCCACATGTGAGATCATGTGGAATTTGTCTTTCTGTGGCTGGCTTATTTCACTTAG
CATATGTCATCCAAATTCATCTGTGTCATAAATGACAAGATATTGTCTTTTCTAT

19346 GAAATAGAGCCCTTTCCCTACCATAACTAAAGATTAACTATATAGCAAAAATACA
TCTTCAAGTAATCATTTTAAATTTTAAATTTTAACTGACAAAATTTGGCATATACATGTTA
TATATATATGTAATGTTGTAATATATATATGTAATGTAACATGATATTTTGTATATGTA
ACCTCTGGAAATGACTAAATCTATCAATGGACATGTTCAATTAACCTCATCTTATCATTTT
TTTGTGGTAAGSACATTTAAATCTACCTCTTAGCAATTTCAAGTATACAAATGTTA
(G,T)
TAACTCAATCACATATTGTACAAATGCATCTCTAACTTATGCTCTCTGACTGAA
ATTTGTATCTTTGACTAAGATCCCTGTAATCCCTCATCTCCACAGCCCTGGTAAC
CACTGTTCTACTCTCTGCTTTCTTTGAGTTTAACTGTTTCAAGTATGAGATCAT
GTGGAATTTGTCTTTCTGTGGCTGGCTTATTTCACTTAGCATATGTCATCCAAATTCAT
CTCTGTTGTCATAAATGACAAGATATTGTCTTTCTATGGCTAAATTTAGTCCATTTGT

20845 TGTACTGGAACTTTGTAGATCAGTTGACAATAATGTGTGGGTGATTTCTGGACTCT
TTATCTGTTTATTTAGTTTATATGTCCTTTTGTAGAGCTCTATGCTTTTGGTGA
CTAGAGCTCTGTAGTCAATTTGAGATCAGGTAGTATGACTCCAGCTTTGCTCTTTT
TGCTCAAAATTTGCTTTGGCTATTTGAGTTTATTTCCATACCAATTTAGGGCTTTT
TTTTTTTGGATTACTGTGAATAATGCCATTTGGAATTTTGTAGGAGATTGCATTGAATCT
(-,T)
TGGGTAGTATGATATTTTAACTGATTTAATGCTTCCAATTAATGAACACAGGGTATTT
GCAATTTGTGTTTCTTCAATTTCTTTCAACAGTTTCTTTCTTAAATTTAATGTTTAA
TTTCCATAGGGTTTGGGTAAAGGTGGTGTGGTATGAGTAAGTCTTTAGTGGTGAT
TTGTGAGATTTTGTATGCCCATCACCTAAGCAGTATACACTGTACCCAAATTTGTAGTCT
TGTATCCCTCACCTCCCTCCCACTTTCCCAAGTCCCAAGTCCATTTGTATCATTC

20845 TGTACTGGAACTTTGTAGATCAGTTGACAATAATGTGTGGGTGATTTCTGGACTCT
TTATCTGTTTATTTAGTTTATATGTCCTTTTGTAGAGCTCTATGCTTTTGGTGA
CTAGAGCTCTGTAGTCAATTTGAGATCAGGTAGTATGACTCCAGCTTTGCTCTTTT
TGCTCAAAATTTGCTTTGGCTATTTGAGTTTATTTCCATACCAATTTAGGGCTTTT
TTTTTTTGGATTACTGTGAATAATGCCATTTGGAATTTTGTAGGAGATTGCATTGAATCT
(T,C)
TGGGTAGTATGATATTTTAACTGATTTAATGCTTCCAATTAATGAACACAGGGTATTT
GCAATTTGTGTTTCTTCAATTTCTTTCAACAGTTTCTTTCTTAAATTTAATGTTTAA
TTTCCATAGGGTTTGGGTAAAGGTGGTGTGGTATGAGTAAGTCTTTAGTGGTGAT
TTGTGAGATTTTGTATGCCCATCACCTAAGCAGTATACACTGTACCCAAATTTGTAGTCT
TGTATCCCTCACCTCCCTCCCACTTTCCCAAGTCCCAAGTCCATTTGTATCATTC

22234 AGAACTTTTATGTTTAAATTAAGTCCACCTATTTATCTTTTCTGTTGTTGTTTGTG
GGGTGTTTGTGTTTGGCTTGGTTTGCATCTGCTTTGGGTCTTGGTCATGAAGTCTT
TGCCTAAGCCAAATATCTAGAAGGGTTTCTGATGTTCTAGAAATTTTATGTTTCAAGTC
TTAGATTTAAGTCTTGTATCCATCTTGAATGATTTTGTATAAGGTGAGAGATGAGGAT
CCAGTTTCAATGCTTCTAGATGTTGGCTTGGCAATTTATCCAGTACAATTTGTTGAATAGGG
(T,C)
TAATATTTAAAGCTTTATATATTTAGGTGTTCTATTTTGGGTACATTTTATTTACAAC
TATCATATCTCTGATGATGATGACCCCTTTCTCATATATATATGTTCTTCTCTCTT
TTTACAGTTTGTGTTTAAAGCTTAAATTTGTCTGATATAAGTTTCACTACCTTTGCTCTC
TTTTGGTTTCTATTTGATGGAATATTTTCCAAAGCTTGGCAATTCATCTATGTTG
TTCTTAAAGATGAAATGAGATGCTGTAGGGCATATGCTTGGGTCTTGTTTATTCATTC

22234 AGAACTTTTATGTTTAAATTAAGTCCACCTATTTATCTTTTCTGTTGTTGTTTGTG
GGGTGTTTGTGTTTGGCTTGGTTTGCATCTGCTTTGGGTCTTGGTCATGAAGTCTT

FIGURE 3, page 16 of 19

WO 02/34922

21/23

PCT/US01/42528

TGCTAAGCCAAATATCTAGAAGGGTTTTCTGATGTTCTAGAAATTTTATGGTTCCAGGTC
 TTAGATTTAAGTCCCTTGATCCATCTTGAGTTGATTTTGTATAGGTGAGAGATGAGGAT
 CCAGTTTCATGCTTCTACATGTGGCTTCCCAATATCCAGTACAAATTTGTTGAATAGGG
 [G, T]
 TAATATTTAAGCTTTATATATTTAGGTGTTCTATTTGGGTACATAATTTATTTACAA
 TATCATATCTCTGATGGATTGACCCCTTTCTCATTTATATAGGTCTTCTTGTCTCTT
 TTTACAGTTTCTGCTTAAAGGCTAATTTGCTGATAAAGTTCCAGTACCTTTGCTCTC
 TTTGGTTTCTATTTGCATGGAAATTTTTCCAACCCCTTGGCATTCACCTCTATGTGTG
 TTCTTAAAGATGAATGAGATGCTGTAGGGGCAATGCTTGGGTCTTGTTTTATTCATTCT
 22247 TTTAATTAAGTCCACCTATTTATCTTTTGGTTGTTGTTGTTTGGGGTTGTTTGT
 TTGGCTTGGTTTGCATCTGCTTTGGGTTCTTGGTCATGAAGTCTTTGGCTAAGCCAA
 ATCTAGAAAGGGTTTCTGATGTTCTAGAAATTTTATGGTTCCAGGTTTATGATTTAAGTC
 CTGATCCATCTTGAGTTGATTTTGTATAAGGTGAGAGATGAGGATCCAGTTTCTGCT
 TCTACATGTGGCTTCCCAATTTATCCAGTACAAATTTGTTGAATAGGGTTAATTTAAG
 [C, T]
 TTTATATATTTAGGTGTTCTATTTGGGTACATAATTTTACAACTATCATATCCTCC
 TGAATGATGACCCCTTTCTCATTTATATAGGTCTTCTTGTCTCTTTTACAGTTTGTG
 TCTTAAAGGCTAATTTGCTGATAAAGTTCCAGTACCTTTGCTCTCTTTTGGTTTCTAT
 TTGCATGGAAATTTTTCCAACCCCTTCCATCTATGTGTGTTTAAAGATGA
 AATCAGATGCTGTAGGGGCAATGCTTGGGTCTTGTTTATTCATTCAITTCAGCCACCT
 22334 GTTCTGGTCAAGGCTTTGGCTAAGCCAAATATCTAGAAGGGTTTTCTGATGTTCTA
 GAATTTTATGGTTCCAGGCTTATGATTTAAGTCCCTTGATCCATCTTGAGTTGATTTTGT
 ATAGGTGAGAGATGAGGATCCAGTTTCAATGCTTCTACATGCTGGCTTCCCAATTTCCCA
 TACAAATTTGTTGAATAGGGTTAATTTAAGGCTTTATATTTAGGTGTTCTATTTT
 GGGTACATATTTTACAACTATCATATCCTCTGATGGATTGACCCCTTTCTCATTTAT
 [A, G]
 TAATGGTCTTCTGCTCTTTTACAGTTTGTGCTTAAAGGCTAATTTGCTGATAA
 GTTCAGCTACCTTTGCTCTCTTTTGGTTTCTATTTGCAATGGAATATTTTTCCAACCC
 TCCGATTCACCTATGCTGTGTTCTTAAAGATGAATGAGATGCTGTAGGGGCAATGCTT
 GGGTCTGTTTATTCATTCAITTCAGCCACCTTTGATTAAGAAATTAATTCATTGTT
 AATCAAGTAAATTTACAGACCAAGGACTTACTACTGCCATTTGTTAATTTGTTTCTT
 23033 ATCTTTGTTGCTCTACTATAGGTTTCTGCTTTGTTGTTACCATGAGGGTTACATAAAGC
 ATAGTTATAAAGGCTATTTTAACTGATACAGCTTAACTTTCAACACTTAAAAAACT
 ATACACTTTTACTCTACCAACTGCOCTOCAATTTATGCTTTGATGTCATAATTTAOCCTA
 GTTTGAGATGCTCCOCTTATGCTATCCCTTAAACAAATTTATGAGCAACAGTCAT
 TTTAATAGTTTGGCTTTTAACTTTTACTAGAGATAGAAATTAATTAACATACCACTAC
 [T, -]
 ACATTTATAGGGTATTCFAAATGACTATGTTTACCTTTATCAGTGAGATTTTGTGTT
 TCAATTTTATGTTGTTAATTAGTATTTCTTCACTTTCACTTGGAGAAATTCACATTAAGCA
 TTTTGTAAAGATGGGCTAGTAGTGGTGAACACCTTCACTTTTGTATCTGGAGATG
 TCTTTACCTCTGCTTCAATTTGAATATAACTTTGTTCCATGATGAATGGACAAAT
 TGTTTTTTAAATATGCAAGTGCCAGGGTAAGCAGAAATTAATTTTCTTTTCTG
 23036 TTTGTTGCTCTACTATAGGTTTGTGCTTTGTTGTTACCATGAGGGTTACATAAAGCATA
 GTTATAAAGGCTATTTTAACTGATAACAGCTTAACTTTCAACACTTAAAAAACTATA
 CACTTTTACTCTACCAACTGCOCTOCAATTTATGCTTTGATGTCATAATTTACCTAGTT
 TTGGAGATGTTCCOCTTATGTTGATCCCTTAAACAAATTTATGAGCAACAGTCATTTT
 TAATAGTTTGGCTTTTAACTTTTATCTAGAGATAGAAATTAATTAACATACCACTAC
 [-, A]
 TTTATAGGGTATTCFAAATGACTATGTTTACCTTTATCAGTGAGATTTTGTGTTTCA
 ATTTTATGTTGTTAATTAGTATTTCTTCACTTTCACTTGGAGAAATTCACATTAAGCA
 TTTTGAAGATGGGCTAGTAGTGGTGAACACCTTCACTTTTGTATCTGGAGATGCT
 TTACTCTGCTTCAATTTGAATATAACTTTGTTCCATGATGAATGGACAAATTTGT
 TTTTAAATATGCAAGTGCCAGGGTAAGCAGAAATTAATTTTCTTTTCTGAG
 23421 CTTCATTTCAACTTGGAGAAATTCACATTAAGCAATTTTGTAAAGATGGGCTAGTAGTGG
 TGAACACCTTCACTTTGCTTATCTGGAGATGCTTACCTCTGCTTCAATTTGAATA
 TAACTTTGTTTCAATGATTAAGATGAACAAATTTGTTTAAATATGCAAGTGCCAG
 GGTAAAGCAGAAATTAATTTTCTGAGACCGAGTTTCACTCTTGTGCTTCAAGC
 GCTGGAGTGCAGTGGGCAATCTCTCAGCTTACCGCAACCTGCTGCTGCAAGGTTCAAGC
 [A, G]
 ATTCTTCTGCTCAGGCTTCTGAGTAGCTGGGATTAAGGCAATGCAACCACTGCTGG
 CTAAATTTGCAATTTTATGAGAGCGGGTTTCTCATGTTGCTCAGGCTGGTCTTGAAC

FIGURE 3, page 17 of 19

WO 02/34922

22/23

PCT/US01/42528

ACCCGACCTCAGATGATCCGCCACCTAGGCTCCCAAGTGCTGGGATTGCAGGTGTGA
 GCCACTGGGCTGGCCAGAACTTACTCTTATTTATCTGAGCTTGAGGAAAGAAATTC
 AATTAATAATTTACATTAACCTAATGGCCAAAGCTGCATTCAAAATAGTAATCAGAAA
 25582 CCCAAAGGCACATAGCCAGTTGCAGCAAGCTAAGCCCAAGATCCATGTCTCTGGAATCC
 CAGCCGAGGGTCTCTCCATTGTGGGACATCAATTCATAGATAATCTTGTCTGGCTGAG
 TTTGAGACCGAGCTGAAACTTCATGGAAATAGCACCAGCATCTTATCTGAAAGACCAA
 GGGGATCTTTGGCTTCATCATATATATACCCCTTATATAATACACATTTAATAGT
 TAAATAGAGCCTTCAGACCATTAATCTATTTTCCCTTGGAAATCAATGTTAACAGA
 (T, C)
 GCTTATACAATGATTTACAGTTCACTGAACACTTTAAGTACTTTCAATGTGGCCAAAA
 TCCAGACCGAGCCCAATGTGTACATGACATTAAGTGTGAGCAGAGCTAGAACTTGT
 GCGGACGCTCTGAGTCTGGAGCTAGAGTTCTTGGAAACACAGGTTTCTGAGCAGGG
 CTTATAGGAAGCAGAGGGGTCTGTGAGACATATTATCTGATTCAATGTTCTATTAAATC
 ATGTCTTAGGAAGCAAGCCACAGGATTGCTTCTGGCAACAGCTACAGCCTGTACTGT
 26407 CCTCTCAAGACAGAGTCTTGTCTATGTAGCCAGGCTGGAGTACAGTGGGTAACTCTGGC
 TCACTGCAACCTCTGCTCCAGGTTTAAAGCAGTCTCTGCTCAGCCTCCGAGTATGC
 TGGGATTTACAGGTGCACACCAAGCCTGGCAAAATTTTGTATTTTATTAGAGATGGGGT
 TCACCATGTTGGCCAGGCTAGTCTCAAGCTCCTGATCTCGAGACCAAGCCTCTCAGCCT
 CCCAAGCGCTGGGACTACAGCCATGAGCCACTGCACCCAGCCAGTCTGTGTCTTTATA
 (C, A)
 CTAAATGTCTCCAGGAGTGTCTAATAGTCCATTAATAGGTATTTAGGCCAGCAGTGT
 CTGACGCTATATAATCCCAATATTTGTGACCAAGGTGGGAAGTCTGTAAGTTAG
 GATCTGAGACTAGCCTGGGCAACATAGGAGAGCCTGTCTTACAAAAAAGAGAG
 AGAGATAGCCAGGCTAGGTTGTGCTGCTGTATCTGCTTACTTGGGGGACAGAGGCA
 GGAGGATCACTTGAAGTCAAGGTTACGGTTACGGTGAAGCAATGTCACGCCACTGCTC
 26473 CAACCTCTGCTCCAGGTTTAAAGCAGTCTCTGCTCAGCCTCCGAGTATGCTGGAT
 TAGAGCTGCACACCAAGCCTGGCAAAATTTTGTATTTTATTAGAGATGGGGTTTCAACA
 TGTGGCCAGGCTAGTCTCAAGCTCCTGATCTCGAGACCAAGCCTCTCAGCCTCCCAAA
 GGGCTGGGACTACAGCCATGAGCCACTGCACCCAGCCAGTCTGTGCTTTTATACCTAAA
 TGTCTCCAGGAGTGTCTAATAGTCCATTAATAGGTATTTAGGCCAGGCAAGTGGCTGA
 (C, T)
 GCATATAATCCCAATATTTTGTGACCAAGGTGGGAAGTCTGTAAGTTAGGAGTCT
 GAGACTAGCCTGGGCAACATAGGAGAGCCTGTCTTACAAAAAAGAGAGAGAT
 AGCCAGGCTAGGCTGTGCTGCTGTATTTCTGCTACTTGGGGGACTGAGGAGGAGGA
 TCCTTGAGCTCAGAAATTCAGGTTACCGTGAGCAATGTTACGCCCTGCTCTCCAGC
 CTGATTGACAGGCCAGACCTGACTCTAAACAAAAAAGAAAAAATTTAAGTAATTT
 26844 TGGGCAACATAGGAGAGCCTGTCTTTACAAAAAAGAGAGAGATAGCCAGGCAT
 GGTGTTGATGCTTTGATTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCT
 TCAGAACTTCAAGGTTACCGTGAGCAATGTTACGCCACTGCTCTCCAGCCTGATTGACA
 GGCCAGACCTGACTCTAAACAAAAAAGAAAAAATTTAAGTAATTTTCAAAACATA
 GCAGAAATATAGCATGGTTATCAGTTTGAATGACACCAACAGCTACTTAAGATAGA
 (G, A)
 TCATGAATTCAGTAAATTTGTGTGGAAAGCTAAGGTGCCAACCCAGCCGCTCTCT
 TAGGTGCTCTCACTGGTGTCTCATGCTACAGCAGGCAAGCATTTGCCAGGAGCTAGCTC
 TTCCCTTCAAGAACAAAGTCTTGTATAGGAGCAGTAGCCCAACCTTGTCTTCTCTC
 CTGAGTCTCTTTTATTTCCCTCTCTTCTTAGGAGTACCGTGGTTCTTAGTATTTGGGG
 TCTTCAACCAAGCCTGCTGTCTGAAAAACCAAGGTATGATTTCTCTTGTACATAA
 28384 CTTCCAGGGAACCGTAGATCTTGTGCTATTTGAGCCCAAGGATCAGTTAGTTTAC
 AAGGACATCTGATTTCTGTGCATCTTTTGGCCATGCTCAAGAGCAGTCCCAACA
 ATGTAAAGCTACTGCTCATAGGCTCAATGCAGTCCACCTTCAAGCAAGAGAAATATTTT
 ATGAGTAATCTCAACTGCCGCTTGTATAGGGAAGGATCATGTTGGAGCCTCCAGCT
 CAAATTTCTCAGTGAACAAATTAAGTCTAAAGTTCAAAAGTTCAATGGCAATTTGGTGG
 (A, -)
 AAAATATCACTTTACTGTCTACTTCAACTCTTGTACTAGTATTTACTATAGTCAGA
 AGAAACATCAATTTTCAAGTATCACTTTCTTCTCTCTGCTTCTGAGGAATGCAATGG
 GCAGGAGTTTGCATGATTTAGTTAAAGGTAACCAATTCCTTCTGCTCTCACTTCAAG
 AGTCACTCCAGACCCCAAGGCTCTTACTTTTCCCAACCATTTTATCTCAAGGCCAA
 GAATGGGATGATTTGCACCTGAAGAACTCTCTGAATGTTAGATCTCAGGGTACAATGA
 28417 GAGCCCAAGGATCACTTAGTTTACAAAGGACAAAGGATTTCTCTGTCATCTCTTT
 TGGCATGCTCAAAAGCAGTCCCAATGTAAAGTACTGCTCATAGGCTCAATGCACTC

FIGURE 3, page 18 of 19

WO 02/4922

23/23

PCT/US01/42528

CACCTTCAAAGCAAGAGAAATATTTTCATGAGTAACTCCAACTGCGGCTTGTATAGGG
 AAGGCATCATGTTGGAGCTTCCAGCTCAAATTCACAGTGAACAAATTAAGTCTAAAG
 TTCAAAAGTTCAATGGCATTGGTGGAAAAATATCACTTTACTGTGTACTTCAGACTT
 [A, C]
 TTGTACTAGTATTTACTATAGTCAGAAGAAACATCATTTTTCAGTATCACTTTCTT
 CCGTCTTGTCTTCAGGAAGTGCATTGGCCAGGAGTTTGGCATGATTGAGTTAAAGGTAAC
 CATTGCTTTGATTCTGCTCCACTTCAGAGTGACTCCAGACCCACCAGGCTCTTACTTT
 CCCCACCAATTTATCCTCAAGCCCAAGATGGGATGTATTTGCACCTGAAGAAATCTC
 TGAATGTTAGATCTCAGGGTACAATGATTAAGCTACTTTGTCTTTGGAAGTTAAATTA
 29265 TATGCAAGTAATAAGTGCATGTATGCTCACTGTCAAAAATCCCAACACTAGAAAAATCAT
 GTAGAAATAAAATTTTAAATCTCACTTCACTTAGCCGACATTCATGCGCTGACCAATCC
 TACTGCTTTTCCTAAAAACAGAAATATTTGGTGTGCATTCTTCAGACTTTTCTCTATAC
 ATTTATATGTAGAAATGTAGCAATGTATTTGTATAGATGTGATCATTCCTATATTTGTA
 TTGAATTTTTCACCTTAATAAAATTCACCTTATTCCTTATCATTCCTTTATGGTATTC
 [A, G]
 TAAATGAATGTACTATAATTTATTTAACTATTTTCCTTATTTGGGCATTTAAGTTATTT
 TGTCTTTAAACACATGCTTGTCAATGGCAACAAAGCCAAATTCACAAATGGGATCTAA
 TTAACCTAAAGAGCTTCTGCACAGCAAAACAACTACCATCACACTGAATGGGCAGCTA
 CAGAATGGGAGAAATTTTGCACCTACTCATCTGACAAAGGCTTAATATCCAGAAATCT
 ACAATGAATCAAAACAAATGTACAAGAAAAACCAACCCATCAAAAAGTGGGTGAGGA
 29484 GTGATCATTCCTATATTTGTTATTTGATTTTTCCTTAATAAAAAATTCACCTTATTCCTT
 ATCATTCCTTTATGGTATTTCTGTAATGAAATGTACTATAATTTATTTAACTATTTTCTT
 TATTTGGGCATTTAAGTTATTTCTAGTTTAAAAACATGCTTGTCAATGGCAACAAAGCC
 AAAATTCACAAATGGGATCTAATTAACCTAAAGAGCTTCTGCACAGCAAAACAACTACC
 ATCACACTGAATGGGCAGCTACAGAAATGGGAGAAATTTTGCACCTACTCTCTGAC
 [A, G]
 AAGGCCTAATATCCAGAACTACAATGAATCAACAAATGTACAAGAAAAACAAACCC
 CATCAAAAAGTGGGTGAAGGATATGAACAGACACTTCTCAAAAGAAACATTTACGCCAGC
 CAAAGACACATGAAAAAATGGCTATCTGCTCACTGGCCATCAGAGAAATGCAAAATCAAAAC
 CACAATGACATACCATCTCACACCAGTTAGAAATGGCAATCATTAATAAGTCAGGAACAA
 CAGTGTCTGGAGGATGTGGAGAAATAGSAGACTTTTACACTGTTGGTGGCAGGAGAA
 30417 ATTCACTACCTAAAAATCTATATATAAAAAATTCCTCCCTTGAATTCAGATCCTTGA
 GACAACACCCACGCTCTAAACCAAAATTTGTTAACTGACGACGCTCCTCTGTGTGAC
 TTTCCATTTTGTCACTATTTTGTGAGCTGTATACCAATATCCACCCAGTTAAACAAAT
 TTCTTGTCTTTTCTGGTACAAACCCAAATAAATTAACAAATCAATAAAGTAAATTT
 CTAAATAACTCACTTTCTCTATATATCTCTCTTGTCTGAAAAATGGGTAGGTAGT
 [T, -]
 CTTTAAAGCCTGCATGATAAATTTGTACTGAATACAAATTTCAAGGTCTGGACATACTAGG
 TATAATTTCTGTGTCTCTGGGCTCTTACCTATTTGGGGTCAAAATAAACAAGTTTATTA
 AGCTTATTAATTTCAATTTCAATATCTTCTTAAACAAATATGTTCCCTGGTGTGTA
 TGGCAATAATTTATTTGTAGGTGCCAGGTGCTTCTAAACTTCTGTATTTTTCATA
 TCCAAATTTACTTTAAATATTTTAGAAAAAGGCTCTGTTAAATTTCTTAATATTTA
 30783 TTTCTGTGTCTCTGGGCTCTTAATTTTGGGGTCAAAATAAACAAAGTTTATTAAGCTT
 ATTAATATCAATTTCAATATCTTCTTAAACAAATATGTTCCCTGGTGTTCATTTGCCA
 ATAAATTTATTTGTGAGGTGGCAGGTGCTTCTAACTTCTGTATTTTTCATATCCAA
 TTTTACTTTAAATATTTTAGAAAAAGGCTCTGTTAAATTTCTAATTAATTAATTA
 TTGTTTTTCACTGACATTTGTGAATTAAGAACCTTAAAAATATGAATCATTTTTC
 [C, G]
 AATATGTGCCACAGCAATTTTGTAAATAAGAAACAGAAACAGGGCATTTATCAAGAG
 ATAAATATCAATATACCTTATATTTCTGTACACACTTTTATACCAACTGTGCCAAAA
 TTGTATATCATATAATGATAACAAAGTTCACAAAGGCTTCTTATTCCTTAACTCTCA
 AATTAGAACTTTCAATAGGTAGGAAGTAGGGGAGCATATATTCCTTTGAAGGTGCAA
 GAAATGTCAATGGCATTCACCATGGTACTCTCAAGCTTAAAAAAATGGAGTGCAAA

FIGURE 3, page 19 of 19

WO 02/34922

PCT/US01/42528

SEQUENCE LISTING

<110> PE CORPORATION (NY)

<120> ISOLATED HUMAN DRUG-METABOLIZING
PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
DRUG-METABOLIZING PROTEINS,
AND USES THEREOF

<130> CLOC0897PCT

<140> TO BE ASSIGNED

<141> 2001-10-05

<150> 60/241,745

<151> 2000-10-20

<150> 09/139,436

<151> 2000-12-19

<150> 09/818,647

<151> 2001-03-28

<150> 09/852,067

<151> 2001-05-10

<160> 4

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 2327

<212> DNA

<213> Human

<400> 1

```
cgccgctgac tctctctccc aggcctgagc tgccctctcc actgaccttc cttcttcccg 60
cgagtcagaa gcttcgagag ggcccagaga ggoggtgggg tgggagaccc tacgccagct 120
ccggggcgga gaaagccac cctctccgc gccccaggaa accgcccggc ttccgagctg 180
cgccagagcca tggattctc ctggctggag acgagctggg cgcggccctt ttacctggcg 240
ttcgtgttct gctggccctt gggtctgctg caggctatta agctgtacct gcgagggcag 300
cggtctgctg gggacctgag ccccttccca gcgcccccca cccactggtt ccttggggac 360
cagaagttta ttcaggatga taacatggag aagcttgagg aaattattga aaataacct 420
cgtgaccttc cttctggtt tggcccttt caggcatttt tctgtatcta tgacctagac 480
tatgcaaga cacttctgag cagaacagat cccaagtccc ggtacctgca gaattctca 540
cctccacttc ttggaagag actagcggct ctgacggac ccaagtgtt ccagcatcgt 600
cgctactaa ctcttgatt ccattttaac atcctgaaag catacattga ggtgatggct 660
cattctgtga aaatgatgct ggataagtgg gagaagattt gcagcactca ggacacaaac 720
gtgggggtct atgagcacot caactcgtg tctctggata taatcatgaa atgcgcttcc 780
agcaaggaga ccaactgcca gacaaacagc acccatgalc cttatgcaaa agccatattt 840
gaactcagca aaatcatatt tcaccgcttg tacagtttgt tgatcacag tgacataatt 900
ttcaaaactca gccctcaggg ctaccgcttc cagaagttaa gccgagtgtt gaatcagtac 960
acagatacaa taatccagga aagaagaaa tccctccagg ctggggtaaa gcaggataac 1020
actccgaaga ggaagtacca ggattttctg gatattgtcc ttctgccaa ggatgaaggt 1080
ggtagcagct tctcagatat tgatgtacac tctgaagtga gcacattcct gttggcagga 1140
catgacacct tggcagcaag catctcttgg atcctttact gcctggtctc gaacctgtag 1200
```

WO 02/34922

PCT/US01/42528

catcaagaga gatccggga ggaggtcagg ggcacccagg gggatgggtc ttctatcaat
1260
tgggaccagc tgggtgagat gtcgtacacc acaatgtgca tcaaggagac gtgccgattg
1320
attccctgca tcccgcccat ttccagagat ctccagcagg cacttacctt cccagatgga
1380
tgcacattgc ctgcagggat caccgtggtt cttagtattt ggggtcttca ccccaaccc
1440
gctgctgtct ggaaaaaccc aaaggctctt gacccttga ggttctctca ggagaattct
1500
gatcagagac acccctatgc ctacttacca ttctcagctg gatcaaggaa ctgcattggg
1560
caggagattg ccatgattga gttaaaggta accattgcct tgattctgct ccacttcaga
1620
gtgactccag accccaccag gctcttact ttocccaacc attttatcct caagcccaag
1680
aatgggaggt atttgcacct gaagaaactc tctgaatgtt agatctcagg gtacaatgat
1740
taaacgtact ttgtttttcg aagtttaatt taccagtaat gatccaagca gctagaaggg
1800
gatcaatgta tgggtggagg attggaggtt ggtgggtag ggtctctgt gaagagatcc
1860
aaaatcattt ctaggatcac agtgtgtcag ctgactctgt ttctatataa ctttgggaga
1920
ttttcagatc ttttctgta aactttcact actattaatg ctgtatacac caatagactt
1980
tcctatatat tctgttcttt tttaaatagt ttccagatc atgcaagtaa taagtgcctg
2040
tatgctcact gtcaaaaatt cccaacacta gaaatcctg tagaataaaa attttcact
2100
tcacttact tagccgacat tccatgcctt gaccaatcct actgcttttc ctaaaaacag
2160
aataatttgg tgggtattct ttccagcttt ttctataca ttttatatgt agaatgtag
2220
caatgtatit gtatagatgt gatcattcct atattgttat tgattttttt cacttcaata
2280
aaattcacct ttttcttaa aaaaaaaaaa aaaaaaaaaa aaaaaaa
2327

<210> 2
<211> 510
<212> PRT
<213> Human

<400> 2
Met Glu Phe Ser Trp Leu Glu Thr Arg Trp Ala Arg Pro Phe Tyr Leu
1 5 10 15
Ala Phe Val Phe Cys Leu Ala Leu Gly Leu Leu Gln Ala Ile Lys Leu
20 25 30
Tyr Leu Arg Arg Gln Arg Leu Leu Arg Asp Leu Arg Pro Phe Pro Ala
35 40 45
Pro Pro Thr His Trp Phe Leu Gly His Gln Lys Phe Ile Gln Asp Asp
50 55 60
Asn Met Glu Lys Leu Glu Glu Ile Ile Glu Lys Tyr Pro Arg Ala Phe
65 70 75 80
Pro Phe Trp Ile Gly Pro Phe Gln Ala Phe Phe Cys Ile Tyr Asp Pro
85 90 95
Asp Tyr Ala Lys Thr Leu Leu Ser Arg Thr Asp Pro Lys Ser Arg Tyr
100 105 110
Leu Gln Lys Phe Ser Pro Pro Leu Leu Gly Lys Gly Leu Ala Ala Leu
115 120 125

WO 02/34922

PCT/US01/42528

```

Asp Gly Pro Lys Trp Phe Gln His Arg Arg Leu Leu Thr Pro Gly Phe
130      135      140
His Phe Asn Ile Leu Lys Ala Tyr Ile Glu Val Met Ala His Ser Val
145      150      155
Lys Met Met Leu Asp Lys Trp Glu Lys Ile Cys Ser Thr Gln Asp Thr
165      170      175
Ser Val Glu Val Tyr Glu His Ile Asn Ser Met Ser Leu Asp Ile Ile
180      185      190
Met Lys Cys Ala Phe Ser Lys Glu Thr Asn Cys Gln Thr Asn Ser Thr
195      200      205
His Asp Pro Tyr Ala Lys Ala Ile Phe Glu Leu Ser Lys Ile Ile Phe
210      215      220
His Arg Leu Tyr Ser Leu Leu Tyr His Ser Asp Ile Ile Phe Lys Leu
225      230      235
Ser Pro Gln Gly Tyr Arg Phe Gln Lys Leu Ser Arg Val Leu Asn Gln
245      250      255
Tyr Thr Asp Thr Ile Ile Gln Glu Arg Lys Lys Ser Leu Gln Ala Gly
260      265      270
Val Lys Gln Asp Asn Thr Pro Lys Arg Lys Tyr Gln Asp Phe Leu Asp
275      280      285
Ile Val Leu Ser Ala Lys Asp Glu Ser Gly Ser Ser Phe Ser Asp Ile
290      295      300
Asp Val His Ser Glu Val Ser Thr Phe Leu Leu Ala Gly His Asp Thr
305      310      315
Leu Ala Ala Ser Ile Ser Trp Ile Leu Tyr Cys Leu Ala Leu Asn Pro
325      330      335
Glu His Gln Glu Arg Cys Arg Glu Glu Val Arg Gly Ile Leu Gly Asp
340      345      350
Gly Ser Ser Ile Thr Trp Asp Gln Leu Gly Glu Met Ser Tyr Thr Thr
355      360      365
Met Cys Ile Lys Glu Thr Cys Arg Leu Ile Pro Ala Val Pro Ser Ile
370      375      380
Ser Arg Asp Leu Ser Lys Pro Leu Thr Phe Pro Asp Gly Cys Thr Leu
385      390      395
Pro Ala Gly Ile Thr Val Val Leu Ser Ile Trp Gly Leu His His Asn
405      410      415
Pro Ala Ala Val Trp Lys Asn Pro Lys Val Phe Asp Pro Leu Arg Phe
420      425      430
Ser Gln Glu Asn Ser Asp Gln Arg His Pro Tyr Ala Tyr Leu Pro Phe
435      440      445
Ser Ala Gly Ser Arg Asn Cys Ile Gly Gln Gln Phe Ala Met Ile Gln
450      455      460
Leu Lys Val Thr Ile Ala Leu Ile Leu Leu His Phe Arg Val Thr Pro
465      470      475
Asp Pro Thr Arg Pro Leu Thr Phe Pro Asn His Phe Ile Leu Lys Pro
485      490      495
Lys Asn Gly Met Tyr Leu His Leu Lys Lys Leu Ser Glu Cys
500      505      510

```

<210> 3
 <211> 31208
 <212> DNA
 <213> Human

<220>
 <221> misc_feature
 <222> (1)...(31208)
 <223> n = A,T,C or G

<400> 3

WO 02/34922

PCT/US01/42528

ccagcctctc ttgggtcctt aaatatagt caaaaagttc cagagttcct ttgttccca 60
tgaaagcaca tggaaaggtg ctggacaggg gcaactggcc ctggagcaga ggaagtaactg 120
catagaactg tccaaagctc agaggaggtc acaccaccag caagaacctg ggtgggagt 180
ggtgagccaa ggggttccca ggctctgacc ctgccaagag aactcattag aaggtcacc 240
accacacata ctatttctcg gtctcatgaa gaaccacggg accgaccag gcaagatata 300
acaaagctga agtttcagct ctggggcaga gcatggatct gaggctttg gccctaccac 360
catgcgatca tatgagggcc atcatcacc catcatgatt tgggggagga ataggccata 420
gaggaatcat atgaaagctt gaaatgccat gaggtaacca gaagaagctg tgaagccag 480
aggattctga gaccctgtca aataacaaca tctagttgaa ggttggagt aggtaggag 540
tggggaagtc tgggaaagaa ggaagctgaa cactgtctgt gtctggctta atggaacatg 600
caaggggcca ggaagaaactt ggtccagatg aagtcaccac cccctggggc ctgtcttttt 660
tttttttttt tttttttttt tgagacggag tctcactctg taccagggtt ggaagtcagt 720
ggcgcatctc cggctcactg caatctttgc ctctcgggtt caagcgattc tctgtcctca 780
gcctcctgag tagctgggat tacagggcgg cgccaccacg cccagctaat tttagtaactg 840
ttagtagaga tgggggttcc ccatcttggc caggatggtc ttgatccctt gccctcgtga 900
tccggccggc tccggctccc aaattgctgg gattacaggg gtgagccacc gcgcccggcc 960
ccctggagcc tgtcttaatc acttcccgcc caaataaat ctggctccag agagtggagc 1020
gtaggcttaa ggaattgggg gcggaagggc ggggaaggtg ggggagggac agtgatagg 1080
agaacaggga attgtagcag aaattgggtt tattgttcag agctgtcaat gaacactta 1140
catatgcttg tcttagcctc aatcaatgaa taatgaatg aatcaataaa tgaatgaat 1200
gtgggcaatg cctataaaga ttgctgggac agggaggtgg ggggagacac cagcttgga 1260
agtcaggcct gttagatcct agttcaccac ctgatacgtt acaataacta aaaccatcac 1320
tttcaaatat tttttactac atttccctgt tatctgtact cgaattttat tatgtttctg 1380
gcctctagag tccagccctc atgggcatga gaaccaagca gccacacagag gctctgaacc 1440
cagaagagca tatgctcggt ttaattggtc gtcatcttag aactgttaat aaagtlttta 1500
tcccgcattt tcattttgca ctgagattca taatttatat agcagggcct gactgtacct 1560
gtatagtggg attactatat gatggtaacg tactgtgcat atcttcccg ttcagtgttc 1620
agtgcctctg tatcgagcag ttgaactagc tcatggtaac cgtcgggaat cagggtgga 1680
atcagttgta aaccatttac cggaaacca ctaggcaggc caccagataa aggaataatg 1740
atggtaacac tccccttacc tctaccctct gggaaatttg gtagaatgcc agaattgaaa 1800
agaaatcttc ttgcategcc atttataatt tgtgataagg aagaaanaac atgacctcag 1860
ctttagcatt attttacaat ataaattcag atcccgtagc tgaanaactgt tggactaaa 1920
agaggaagct ccaggagcgc aaagcagtt gggccgaacg aagcgtgagc gctttggtaa 1980
cgggctagaa atcccgacg cgcgcctgcc tctctctccc aggcctgagc tgcctctccc 2040
actgccttcc ctcttctccg cgaagtcagaa gcttcggcgg gggccagaga ggcggtgggg 2100
gtgggagacc ctacggcagc tccggggcgg agaaagccca cctctctccg cggcccatga 2160
aaacggcggc gttcggcgtc gcgcagagcc atggaattct cctggctgga gacgcgctgg 2220
gcgcggcctt ttaccctggc gticgtgttc tgcctggccc tggggtgcl gcaaggccctt 2280

WO 02/34922

PCT/US01/42528

aagctgtacc tgcggaggca gcggtgctg cgggacctgc gcccttccc agcgcctccc
2340
acctactggt tccctgggca ccagaggtta aatggaagg gaaaaggnta gaaaaggagg
2400
aagagggggg cggaggagga tgcggcagag gagccagcc ggcagagaga cgcagctttc
2460
ttccatccct ggggacctc cggcttgca cggcctttcc agcccgccct gtggtcttca
2520
gcacatttt tccctgctct ggagaattgc ttcccgag cccacaggg aagggtcaca
2580
aaagaggag ctttggggc tgggagagag ctatttaag aacctgaata tggaaaaaga
2640
aagcgagctg taactcaagt ctgtcttca ttgttccac aagccttcca catgtgttc
2700
tttaaaata gcctgttatt ctaataact tattagttgc agaaatatg caaatctat
2760
cccaactggt ggcacctta gtccattta acaagagaaa attttttt cctaagattc
2820
ttgtgaagta aggagcagcc ccagccagcc actcgagaaa tactgattga tggaaatttg
2880
taaaggagga ctgttagctt ttggtcttc ccgttttta aatccactcc caccctaact
2940
taagggtttt attcattca cgcactctga gtggcaattg tgtgataggt actaagatta
3000
caaaagagag ctaagtccct cccctgacc aaccaagta ggtgcagact taggccacag
3060
agagaaatg aaaaatttaag gcaatgggtg ctttactaga ggcctagaga caagggaata
3120
tctgtcggg gaaagtata atctccgct agagaaggaa ggaagctctg tgaagggtg
3180
agcagctct taaaggatg ttgggtggg tggggaagg attccagcag agctactaca
3240
cgatccctt gttcccccac ttctagtct tcttatata aagcaaccac tttaactct
3300
tttatcggt tcttctgga tttaatact tttttgtaa atagtattac catattgcat
3360
ctattaatt aataagttta gacatctgct gtgggtttag tatggtttgt tctcccccac
3420
caagcctcct gttgaaattt gattcccaat gttggagggt ggtctgctg ggagatctt
3480
gggtcattg gatggatccc tcatgaatgt cttggtgag ctgtctcct cataagttct
3540
cactctctta gtccctcttc aaccocagc actgattgtt gaaaagagcc tgcacctcc
3600
tcccctctct ctctctgtct ctccactgt ggtctctgca cacaactgct cctgttccct
3660
tccactatga gtggaagcug tctgagatcc tccgagatg cagatgccaa tgcctgctt
3720
cttgtacagc ctgcagaatt gtaacccaaa taactctctt tgtgaatgac ccagcctcag
3780
gtattccttt acagcaaac aatgtlacta agacaacac cactatgaa cttctttatg
3840
acaggcaatc acttaccctt catatlccac tgtccagta actatatagt attgtattt
3900
ttaaatagaa aaacttctat ttgtattatt ttattatgc aatgttatt tactgtgat
3960
ctaaatggc ctcttctatt ttatttctt ttctcataga actttttccc cacccccaca
4020
gtattggnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
4080

WO 02/34922

PCT/US01/42528

nnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn
4140
nnnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn
4200
nnnnnnnnnnnn nnnnnnnnnnn ntgttatgta tctctactgt ctcatgaata ctatgtcgtc
4260
tggtgtttta attgaattgt ttggcatcc ttgtcaaaa tcaattgaco ataatgtca
4320
aggctctatt ctgagctctt aattctaate cattgatcta tatgtctatc ctactcatg
4380
gacacagaga gtagaaggat ggttaccaaa ggctgggaag gatagagggg agctggggga
4440
ggaggtaggg aaggtaatg ggtacaaaa aatagaaag aatgaataac acctactatt
4500
tgatagcata gcagggtggc tatagtcaat aataactgta cacttttaaa taagagtggt
4560
aataggattg ttgtcaactc aatggataaa tgcttgaggg gatgggtacc ccattcttca
4620
tgatgtgctt atttcacatt gcatgctgt atcaaaaaca tctcatttac tccataata
4680
tatcacctta ctatgtatcc acaagtatta aaattataa ataaataaat tatatagcta
4740
tccttatgct agtaccacac tgccttactg ttgctttgta gtaagctttg aatcaggaa
4800
gtatgagtc cccgcacttt ggtattttcc aagattattt tggtctttg gaatccttga
4860
tttctataca aattttagac tcagcctatc aattttatac aggaaccag ctagggttct
4920
gcttgggatt gcactgaatc tgtatgacg ttggggatt attgccatct taagaatact
4980
aggctctctg atccatgaac acagaaagcc ttccgttta gttaggtcat ctttaattct
5040
ttttgttgtt ttttttgtt ttttgagaca gagtctgtct ctgtcgccca ggctggagtg
5100
cagtgaagca atctcggtc actgcaacct ccgctctctg gattcaagcg attctcctgc
5160
ctcagcctcc caagcagctg ggaactacag cacatgcac cacaccaact aatttttga
5220
ttttcagtag agacgggggt tcaaatatt ggccaggcta gtctcgaact cctgacctg
5280
tgatccacc gccctaccct cccaagtgct tgggattaca ggctgagcc accactccg
5340
gctttcttta attttttta acgatgttt tgtattttc aaagtataca tcttgcatt
5400
cttttgttaa atttatttgt ttgtttttt ttaatttcac ttcagactat ttattgcatt
5460
catagtgtt tagagtccac attccctctt gactgtcact aagttttttt tttctgttt
5520
ttgagagggt tctatcagaa ttttgagat cagagatgac ggaatgtca aactgtctaa
5580
tattccaac cctcccatc tatcagata ggaacctttt ggtgattcac catgcaggga
5640
aatctagtat ctaaggctca aaagtgata ctgttttaca taggcagtaa cattttattg
5700
ctacataata actacatatt tatggagtac ctgtgatatt ttgatactg catcaatgt
5760
gcagtgatca aatcagggtg tttaggtat tcatcacttc taacatttat tatttatgt
5820
tgtttggaac atttcaagtc tcttcaagct cticagaat attcaataca ttattgttaa
5880

WO 02/34922

PCT/US01/42528

cagtgctatt gaacactgga acttattcct tctatctaaa gacagtaaca ttttaagtat
5940
agtcataagg ttacagaagg ataaagtgtg tatagggaaa attccctaca agatgagaa
6000
ttcattccctt actcttagta atacaggtct tcaaocatgc caaggatatt cctcccttgg
6060
agctttgaac atgcacgtct gtggttatat tgctctccct gcaaatattt cctaaagag
6120
gcttgccctg accattcaga ctaaaatagc acccttagta ctctctatct ccaacctat
6180
tattattatc ttggccctta tcaactctctg acactatact gtatactctt ttgcttgctc
6240
gtttattatc caccactaac tacatataaa aatctgtgag aggtaggatc ttigtgttgc
6300
actataaacc tagtgcatgg tccagttcct ggtgcataat aggtgctcaa taatccctt
6360
gttgaatgca taatatattt aggtgctgag aaaaatttctt tttcaaga tcaatttact
6420
gcatagaata ggcaggtggg ttgacattt attcaatagc caacatatgg gacctaggat
6480
gtacatatgc aagtgtgtgt gtgtatgtgt gtgtgcatct gcatgtgtac ttggatgtac
6540
tgcaagagaa atctatgtag ctatgtatga taaagcaatt gggctccaga gttaaectgg
6600
agtttgaatc ctcatatgtg gttgccagct gtacacactt gggcaatca tttaaactag
6660
tctgtagggc tcaatttccct catctctaaa gtagggattg taatcatatc tacttcatag
6720
ggttcttgat gtaaatatta aataacatag aacatggaaa gcatttagca gcacctagt
6780
catagcagtg ctgtataaat gttcgtctgt gctatttggg ggcactatgc attttctgae
6840
catttctgaa caatgtttac taaatatatg tagtaccgtt ttcaagtgt atttagatgc
6900
ttctctgggg atgaagaat ataaattaaa tatagtacag tattcacamc agttttctgt
6960
cctttttgtc tagtcagtag ttacaanaag tataatgaag tactttcata tggctgggg
7020
gtttatgaaa attttttacc taacaaaca attgtcatat tagtttaca tattcatgag
7080
ggcaaaaggc ttgtcttccct tatatttctc tgtatctcta ccacctggta cgtgtgatag
7140
acaataaata cttgtgtgtt tattgtttgt aaatgaataa atgaanaaat attcacattg
7200
ttgaaacca ctactctgga tagtcagtgg gtgcttatca ctggcttgat tatggcaaca
7260
ttaacaaaa agtgcatat tttgaaact aggtttcaag actetcaacc tttaagtggc
7320
cttgaaatat ccagagaaca ctttatgggt taaaattgct aaatgataac agagaaaaat
7380
gggagccaga gttgtccacc lctccagagg atgagagcaa acaatctgct agcagatacc
7440
glgtgattgg tcacacgagg aaaaatctgg cagccttaag attactttgc agcgggggac
7500
tcccaccatc atgtcaagt gtgtagatgg gcacacaaa acacacacat gcagggtgcc
7560
tccactttac acaagaagca aatgtaaatg aatcttgttt tcagtgattt agagaacaa
7620
tttaagttag ccatcttca lctgcttcta aaagcaaaa ctccttctct ggtggtagta
7680

WO 02/34922

PCT/US01/42528

taactgggga ttccaggatt agcgtgctaa tcttataagg ctttgaatt tattagactt
11340
tgaactgtt tctcacaata ttaatacat ccatcccaga ggtaagcttc taatttcacc
11400
ttcatctatt aaattgcatt gcacattaat acgagtacta ctttgatact ccactgttgc
11460
atgactgcct gtgggtcatg gttaactcac gctgctgtg ttectcatct atccttcac
11520
tcctctaatt aaatggcata aggtttttctg ccttttattt ctcaaggaaa aggactagcg
11580
gctctagacg gacccaagtg gttccagcat cgtgcctac taactcctgg attccatttt
11640
aacatcctga aagcatacat tgaggtgatg gctcattctg tgaatatgat gctggtaagt
11700
aaagggggga agtgctctgt gcattgcgaa atgctccag caatggacag tattaggtat
11760
gtgtttttgt ggcctgaaa ataaaaatc agtttctaaa attttaacca atgtacacgt
11820
acttattgaa caataggtgt ctgtaaaaaa ttgtttatgt tctttgagtg ataataaa
11880
taaaaaagatc ttgtcctctg tcttagatat attttgagat ttlatggcag caaaccaagt
11940
accataatggg gatagttaga tagtaagtgc tgtagatgtg ttcatggag ggcgggtctg
12000
tacaacccta occcaagtc tgaggaact gagaggtga agaaaaggo tgacagtto
12060
tcaaaaagaa acattcaata gaggccttca acaaaaaacc atctctctctct atctctctctct
12120
atctctctctct atctctctctct atctctctctct atctctctctct atctctctctct atctctctctct
12180
atctctctctct atctctctctct atctctctctct atctctctctct atctctctctct atctctctctct
12240
atctctctctct atctctctctct atctctctctct atctctctctct atctctctctct atctctctctct
12300
atctctctctct atctctctctct atctctctctct atctctctctct atctctctctct atctctctctct
12360
atctctctctct atctctctctct atctctctctct atctctctctct atctctctctct atctctctctct
12420
atctctctctct atctctctctct atctctctctct atctctctctct atctctctctct atctctctctct
12480
atctctctctct atctctctctct atctctctctct atctctctctct atctctctctct atctctctctct
12540
atctctctctct atctctctctct atctctctctct atctctctctct atctctctctct atctctctctct
12600
atctctctctct atctctctctct atctctctctct atctctctctct atctctctctct atctctctctct
12660
atctctctctct atctctctctct atctctctctct atctctctctct atctctctctct atctctctctct
12720
atctctctctct atctctctctct atctctctctct atctctctctct atctctctctct atctctctctct
12780
atctctctctct atctctctctct atctctctctct atctctctctct atctctctctct atctctctctct
12840
atctctctctct atctctctctct atctctctctct atctctctctct atctctctctct atctctctctct
12900
atctctctctct atctctctctct atctctctctct atctctctctct atctctctctct atctctctctct
12960
atctctctctct atctctctctct atctctctctct atctctctctct atctctctctct atctctctctct
13020
atctctctctct atctctctctct atctctctctct atctctctctct atctctctctct atctctctctct
13080

PCT/US01/42528

11

WO 02/34922

PCT/US01/42528

actatcttta ttatggaca ttattattaa tacaatatata agtaggcact taagagttcc
14940
agacatacat ggaatatggc tttttgcaca gcgattgcag taataataat gacaagctaa
15000
aaacattcat gcaacatagg aatggagagt ggaacagagt aaacatggac atgcacccga
15060
aagaatattg attcaaaaac agtttttqca agcataaaca caaaagtga aatagattaa
15120
gcttttttga caattcaaca ttacttgta tgaatgccat aatggagaat acttatcag
15180
cagtgaatta atccttcac agcttcacca ctactagca gttactagta agttacttac
15240
tgctttgttt cagtgtcatc tataaaatgg agattaaaaa agaactatc tcatacattt
15300
gttggttcga tgggtgggtt aatatatata aagcatttag gacagtgcct ggcactgaat
15360
agatgttaaa tgtaaatgat agttatgta aatgtctttg ctccagga ttttgcaaga
15420
cacaccaaca tatgcacact taccatata tatatgcata catgcacata gatattataa
15480
agaggacact cagagaagca ggttatcaac aatttaagga ataagtggc attatcaata
15540
gcagcagttc ccaagtcttt ctgcacatt gcacacacag aaagtgtta tgtttttgtg
15600
cttcattgga gtaaacagga atggatttgg ggaagctat acagaacttt gtaaaaaaa
15660
atctttactt ttaaatatt atacaattat gatgaanaag caaatgcaa agtggtacgg
15720
aaaatattaa atgttcaatt tattcaaac ttaaacctt ttaattttt ttttttttt
15780
ttttttgaga tggagtctct ctactcagg ctggagcgca gtgtgtgat ctgagctac
15840
tacaacctcc acctccagg ttcaggcaat tctctacct cagcctctg agtagctggg
15900
attacaggca ctgccaccac acctggctaa ttttttttaa ttgtttatt ttatttagtc
15960
aaatatatca atattttatt ttattgcac tggattttta gtaatcaaa aaagccattc
16020
tctattccag ggtttctcaa cctcagcac taatggcttc ttagattaga taagtccttg
16080
ttgtcaagat gtgtgcattg taggatgttt agctacatcc ctgacatcta cccactcgat
16140
gtagtagagc tctgatagtt atagcaacca taataaactc cagacattat tgaatgttcc
16200
caggggcccc agttgagaac cctgcccctg taccaggtt gtagagaaaa tttttatgt
16260
tttcttgtag taactgtata atttcattat ttcatattt aaatagaga tctaaactoc
16320
atttagaatt tattcctata tatgggtgga ggtattgac taatttttcc aaatgtttat
16380
ccagttgtcc catcaccatt atttaaaagt ttatcttttc aagtgtttg agataccat
16440
cacattctaa acggatacat gtactggat ctgttttga taagagtata ttggatgtt
16500
ctcgtgtatt ccattgatct atctaccaat gtaccagaat cactctgtt taattcaaga
16560
gattttgtgg cttttttcaa cattoalaga ccttattttt agaaaagttt taggtttgca
16620
gaaaaattca gcagaagta cagagagttc tcatattacc catgtaaaa acctgtacat
16680

WO 02/34922

PCT/US01/42528

gtacccctgt atctaaata aaagtgaat ttttttaaat agtaataaa tattacctc:
16740
gttccatatt tttgttttgt tttttttctc tcagctcctt caattataa tatattggca
16800
tttctttgcc tgtcttctat ttcattccat tttatttaat aucttttccg tgaagataaa
16860
atattagact gaggaagaaa agataattg gtcacttgca tctaaacttg aatcatctt
16920
aattttattg cccacatact gatggaact atgtttttta tttgtgttgt ttatctttgg
16980
agcttttaac aaagtccct ttgatgagaa aataaaccat ctgtgaaat tagatctat:
17040
taaactgtctg gaactcagcc aagatttga gctattcact aacctggct tgccttataa
17100
tttatttgac tttgccatca ctttggtaat tggaaactat ttttctaccc agatacaata
17160
atccaggaaa gaagaatac cctccaggct ggggttaaag aggaatacac tccgaagagg
17220
aagtaccagg attttctgga tattgtcctt tctgccagg taactctttt aattttctaa
17280
gcctgtctca gtgaccagt aattatgtaa gtggtgggt aagtgggaat gggatggga
17340
gacaaagata aaaccgattg actaaattta actgtacttt gaattgatga gcagcttcat
17400
gcaatttgag acaagagag aattctgcaa ctgtgtcgt agaggagggt tagtaagac
17460
taaacgaacg atttgacaag atttgaggat tgcataatgg atacatggat tttagggcat
17520
catgaaaaaa tggtcacatg gataaacgta aaattatga tgataaggtc ctgggaatc
17580
tgggagtttg aagagaattt ctagggcctg ttgatcaggg gcccttttgt caaggcctgc
17640
tttctttatc taaccttgtt tctcctttat gctttgggca gaalatggtt tataccacat
17700
attttgtgaa ctgaattaaa atttaaaccc ctatttaag ctctgctttt tccctcaaa
17760
tcattattgt ggttgtatct ccaaacattt ataaactggc attttattta aatatattgt
17820
attgtacttt ctaggatgaa agtggtagca gcttctcaga tattgatgta cactctgaag
17880
tgagccacatt cctgttgga ggacatgaca ccttggcagc aagcatctcc tggatccttt
17940
actgctggc tctgaacct gagcatcaag agagatgccg gagggaggtc aggggcatcc
18000
tgggggatgg gtcttctato acttggtaag atctgcaccc cttaattttc ctgctagttt
18060
tcccccctgag attttgcttt attttttgct ctggtacctt agtgacccta gtgcctcagg
18120
atatgtgttg gtgaacaga agaagtggc tacttttctg ttctttctaa agagagctcc
18180
aaattattct ctgtctttc aggaataaaa aaaaagtta ttatccata aattgtctgt
18240
cattggtttt ctaatcaatg gtgtgtgaaa tgtcttattt ctttatttca ccttggtct:
18300
gatgcattgg aatgaggac ttgtccctg ggctggcact tagaacttaa acatagggt
18360
ccaagtggag ctctcttct gagagagctg aatgattagc tgcattattt aaggctcatt
18420
ttagacatct ccagccgct tgtaccaat ttattctctc aggatgatt ttagactta
18480

WO 02/34922

PCT/US01/42528

gacataatat tcgatgatat atactatagt caagtttagc aaatatggac tgaggacatt
18540
ttaaatactg agactttttt tatgactaca atttatgttg ggccctgtct tcggtgagct
18600
aatggctcaa tacaggagac aggagacaga cctccaaatt gcagtgtagc ataatgaggg
18660
caatgataga gatatgtgct ggctaacaca aagacataga agacaggtae ctaccctggc
18720
atgggagctc aaggagactt ccttgacatt tacgctgact gcaggataag taggagttag
18780
ccaggtggaa actgtcatct ctatcttgct agactttaag catatactgc tgttaataag
18840
gcccaggtta tgetgtttgc aaagataaaa tgtgttcctg acataatact ggtcaagggg
18900
acagaaagac agaaatgcta aggacaattc agcagcagac cagataaaaa acaacctatt
18960
tcatttgcaa aagtcactc aattgaaca tttgtaaac caaatttgac attataaag
19020
tatatcagag atctcatttt atcaggaaat agaagccctt tcctaccata aactaaagat
19080
ttaatctata tggcaaaaaa taacatgttg agtaatactt ttaatttat ttttaactg
19140
acaaaaattg tgcataaca tgttatatat atatgtatgt gtgtatatat atatgatgta
19200
caecatgata ttttgatata tgtatacact gtggaatgac taatctatc aatggacatg
19260
ttcatttaact catactatc attttttgt ggttaaggaca tttaaatct accctcttag
19320
caattttcaa gtatacaaat tgttagtaac tccaatcaca tattgtacaa tgcattctct
19380
aaacttatgc ctctgtctg actgaaattt tgtatccttt gactaacatc cctgtaatcc
19440
cccattctcc cacagccoct ggttaaccact gttctactcl ctgcttcttl gagtttaag
19500
ttttagattt ccacatgtga gatcatgttg aatttgtctt tctgtgcttg gcttatttca
19560
cttagcataa tgtcatccaa attcatctct gttgtcataa atgacaagat atttgtcttt
19620
tctatggcta attgttagtc cattgtttat atatatacca tgtttctctt atccatttat
19680
ccagtgatgg acacttaagt tgaattctat atctgggcta ttgtgaataa tgcgtcaatg
19740
aacatgggaa tgtagatgtc tcttcaatgc actgatttca ttctgtttgg ttgtatatcc
19800
agaagtggaa ttgctgcac atatggtagt tctattttta attttttgag gaactccgt
19860
acaattttcc atatggctgt actaatlac attccaacca aaagtgtata agggttctgt
19920
tttctccaca tctcaccaa catttgtctt ttgtgtaata accattctaa tgagcatgag
19980
gtgatgtctc attatggttt taatttacgt ttccctgatg attagtgatg ttgagcattg
20040
tttttaaatc ctgctggcca ttcatgtctt ctttgtagga atgttatatt aggtttttct
20100
catttttcaa tctagtattt tgtttcttg cttttgaatt gtgagagttc ctcatatatt
20160
ttgaatatla accccttacc agatgtatca ttgcagaca tgttctccca tcccttaagt
20220
tgtctcttca ctatgttgat tgtttccttt gttgtgcaga agctttttag ttgtctgcaa
20280

WO 02/34922

PCT/US01/42528

aaccatttat ctatcttttc ttctgttgac tatacttcca gagggtgttc caaaaaatca
20340
ttgccaagaa taatatcaag aagcttttct ctatgttttt ttctgttagt ttatagttt
20400
cagggtcatat gtttaaatct ttaattccatt tttagttgat tttgtatat ggagtgagat
20460
aaagggtccac ttttattctt ctactagtgc atatccagtt ttctcaaac catttattga
20520
agatactgcc ctctcaccac tgtatgttac tgaaccttt gtagatcagt tgacaataaa
20580
tgtgtgggtg tatttctgga ctcttctcc tgttttalla gtttatatgt ctctttttt
20640
agaagctcta tgcgttttg gtgactagag ctctgtagtc aatttcagat caggtagtat
20700
gatgcactcc agclitgctc tttttgtca aaattgcttt ggctatttga gtttttlat
20760
tccatagcaa ttttagggct ttttlttt ttctgattct gtgaataag ccatttgaat
20820
tttgatggag attgcatga atctttgggt agtatggata ttctaacagt attaatgct
20880
ccaattaatg aacacagggt attttgcaat ttgtgttttc tcaatttct ttcaccagt
20940
tttttttctt aatttaattg ttttattcc atagggtttg ggtaacagggt ggtgtttgt
21000
tatgagtaag ttcttttagt gtgatttga agattttgat gcacctatca cctaagcagt
21060
atacactgta ccaatttgt agtctgtat cctcaccct cctccacca tttccccaa
21120
gtccccaag tccattgtat cattcttatg cctttgcat ctcatagctt agctccact
21180
tatggtgag aacataaat gtttgggtct ccatcttga gttacttcat ttagaattt
21240
ggctctcaat tccatccaga ttgctgcga tgcctttatt ttgttcttt tcatggctga
21300
gtagtattcc atagtatata catccacca tttctttat cattcttgat tgatgggat
21360
ttggactggt tccatgtctt tacaattgag aattgtgctg ctacaaacat gcagggtcaa
21420
gtgtctttt catataatga ctctcttcc tctgggtaga taccctgtag tgggattgt
21480
ggatcaaatg gtgttctac ttttagttct ttaaggatc tccacactgt tttcattagt
21540
ggttgtacta gtttacattc ccaccaacag tgtagaagt ttccctgttc actgtatcca
21600
caccatcatc tattattatt tgattttttg attatggcca tcttgcagg agtaagggtg
21660
tattgcactg tggltttgat ttgcatttcc ctgatcata gtgatgtga gcatttttt
21720
atatatttgt tggccatttg tacatcttct ttgagaatt gtctattcat gtcctttgt
21780
catttttga tgggattatt tgttttttc ttgctaattt gaggctccctg tagattctg
21840
atattagacc tttgttgat gtgtaggtg tgaagatttt ctccactct ttgggttgc
21900
tgtttactct gctgattatt tcttttctg tgcagaact ttttagtta attaatgccc
21960
acctatttat ctcttctgtg ttgttgttt ttgggttgt tttgttttg ctggttttg
22020
catctgtttt tgggttcttg gtcataagt ctttgcctaa gccaatatct agaaagggtt
22080

WO 02/34922

PCT/US01/42528

tctgtgtgt ctagaatttt tatggttcag giccttagatt taagtccctg atccatcttg
22140
agtgtatttt tgtataaggt gagagatgag gatccagttt catgcttcta catgtggctt
22200
gccaatatc ccagtcacat ttgttgataa gggtaatat ttaagcttt atatatitag
22260
gtgttctct tttgggtaca tatttattta caactatcat atcctcctga tggattgacc
22320
cctttctcat tatataatgg tcttctgtc tctttttaca gtttttgtct taaagcctaa
22380
tttgtctgat aaaagltcag ctaccccttg tctcttttg tttctatttg catggaatat
22440
ttttttcaa ccttcgcat tcaactctat tgtgttctta aagatgaat gggatgctgt
22500
aggggcatat gcttgggtct tgttttatto attcattoag ccaccccttt gatlagagaa
22560
tttaattcat ttgtattcaa ggttaattat gacagacaag gacttactac tggcattttg
22620
ttaattgttt tcttgatgtt ttatagatct tctgttctt tcaacctctc ttactctltt
22680
cctttgtgat taggtgcttt tctctagtgg tgcactttga ttttacttt ttatcttttg
22740
ttgtctact atagggtttt gctttgtggt taccatgagg gttacataaa gcatagttat
22800
aaaagctat ttaaaactga taacagctta actttcaaca cttaaaaaa ctatacctt
22860
tactctacc aactgccctc ctttttatgt ctttgatgtc ataattacc tagttttgga
22920
gatgtgtccc ctatttgtgt atccctaac aaattattgt agcaacagtc atttttaa-
22980
gttttggctt ttaactttat actagagata gaattaatla acataccacc actacattat
23040
tagggatttc taaattgact atgtatttac ctttatcagt gagatttttg tttcaattt
23100
tcattgtgtc aattagtatt ctltcatttc aacttggaga atcacatta gcattttttg
23160
taagatgggt ctatgtatgg tgaacacct caactttgt ttatctggag atgtctttac
23220
ctctgcttca ttttgaata taactttgt tccatgattg aaatggacaa aattgtttt
23280
ttaattatgc aaagtccag ggttaagcaga attactcttt ttttttttt ctgagaccga
23340
gtttcactct tgttgcccag gctggagtg agtggcgcaa tctctcagct taccgcaacc
23400
tctgcctccc aggttcnagc gattcttctg cctcagcctt cctgagtagc tgggattaca
23460
ggcatgcacc accatgctcg gctaattttg catttttagt agagacgggg tttctccatg
23520
ttgttcaggc tggctttgaa cacccgacct cagatgatcc gccacctag gcctcccaa
23580
gtgctgggat tgcaggtgtg agccactgag cctggccaga attactctta tttatcctga
23640
gcttgaggaa gaaagaattc aaatttaaaa tticacatta cctaattggc aaagcctgca
23700
ttcaaaataa gtaatcagaa aaactataa aaacacata agataaacag actaaatata
23760
tgcagtcatt ttatggaacc atctgacta gattggatgc agactaggtc ggtgcgaat
23820
ttaaaaaaa ctttattctt ctccactta taaactttaa acctgctttg tggagcaagt
23880

WO 02/34922

PCT/US01/42528

tctttttatc tctggggaaa gatcctgagc aagtctcata gaggctcat tcatTTaaat
23940
cacaagaaaca atcttaggtc agtaattaaa ctatctggcc cagtgtatc ctgaacttt
24000
caaatactta tccacttgag ctcttcttcc catcccaagt tggtaattct ttggtcctag
24060
aagccagcag tggtttatca tgcacttatt ctactgact agctcccaa taccagtag
24120
ctgctgtttc tggccctcc aggaatggtt ttaggaggaa aggggataag gataaaggg
24180
ctggtactat tgtgatcatg ccaagggtc tgggtggatc tcatgcttc cctttcttc
24240
aagaggaac tccctttctt ggagactctc tcaatagaa tttccagag tgaattcagg
24300
gacaagagaa taattgtcct taggcagact ctttttcaag ctgggtccag agctttccct
24360
cttgccagt ttttgggtta aggacacagt tgcacatcct tgccttgcct ctgctgctgt
24420
cctctgctt tctgtctgt ctgagttata gcctttcaca tcaatcctgt actcccaaa
24480
ctccagggag caaagtcag atcatctaag tgatcctctt gaagcctctt gtttaagatg
24540
gggggaagcac ccttcccttt ccatggcact ctggcattcc acaacactt taataattt
24600
tttctctcaa aattcttaag cctctcctct ttaatcctc gccatttita tglattatta
24660
ctttatatga tgagctaga gttacaaaac tggtttttag aaatctcctt agcaaatgtt
24720
ttactgctag tttagcagct cactttataa taaggatata tgatatattt ctttgggtcc
24780
tctgctctg ggacctcagc tcatcctgag gcagagagtc ccattttac attctgtac
24840
ataaacagc ggcaaatgg cttaacctg agggtaataa ttaccaggaa caaacagaaa
24900
acagaaaaaa agtaactgg ttatgatata tgagtccctt cctccctca tcttcacag
24960
gaccagctgg gtgagatgc gtacaccaca atgtgcatac aggagacgtg ccgattgatt
25020
cctgcagtc cgtccatttc cagagatctc agcaagccac ttaccttccc agatggatgc
25080
acattgctg caggctctta cattcttttc ctaagcagtt cttaagagct atgggtacct
25140
ggagaccaca gtgacaaga ttatgtgagc tcttagcact tggagaagtc aaaaagataa
25200
gctaacatgt gacttaggtt ttatcaccta tgaggagctc agaggataat gctttggtca
25260
gacatgaatt tcaatgactt tcccaaggc acatagccag ttgcagcaaa gctaagccca
25320
gaatccatgt ctctggaac cagcccaag gtctcttcca ttgtgggaca tcatctctaa
25380
gataatcttt gtttggctga gtttgagacc gagctgaac ttcatggaaa atagcaccag
25440
catctttatc tgaagacca agggggatct ttggcctcat catcataata tccctctat
25500
aaatalacaa catttaatag ttaatataga gccttcagac ccattctctc atttttccc
25560
ttggatcca atgttaacag atgcttatac aatgatttac agttcactga acacttttaa
25620
gtactttcaa tgtggcccaa aatccagag cagcccaat gtagagatga cattaatga
25680

WO 02/34922

PCT/US01/42528

tgtagacaga gctagaactt gtgcggagac cctgagcttg gagcctagag ttcttcggaa
25740
caacacaggt ttctgagcag ggcttatagc aagcagaggg gtcctgtgag acatattatc
25800
tgattcaatg ttctattaat tcatgtctta ggaagcaagc caacaggatt gcttctggca
25860
aacacctaca gctgtttact gtaactttgc tgacagacc agaatattt tctggaagct
25920
agaattattt ctggaacca aataaccctc acattctctc tctttgttt tgtactctgt
25980
ttctcccaa accacatgga tatttgcaa aattctccac ttccatag tgatagcac
26040
caatggaaat ttgtcatggg atctgcatga cagaatcaca gtctgtgtg tgtgtgtgtg
26100
cgttttctc tcaagacaga gtcttctat gtagccagg ctggagtaca gtggcgtaat
26160
ctcggctcac tgcaacctct gctccagg tttaagcagt tctctgctc cagcctccg
26220
agttagctgg attacaggtg cacaccacgc ctggcaaat tctgtattt tattagagat
26280
ggggtttcac catgttgcc aggttagctt caagctcctg atctcgagac cagcctcct
26340
cagcctccca aagcgtggg actacagcca tgagccactg caccagcca gttctgtct
26400
tttatcccta aattgtctc aggagtgtt aatagtcct taataggat tttagccagg
26460
cacagtggct gacgatata atccaatat ttgtgacac caagtgagg agactgctg
26520
aagttaggag tctgagacta gctgggcaa catagggaga cctgtcttt acaaaaaaa
26580
aaaagagaga gatagccagg calgggttg catgttgtt ttctgcta cttgggggac
26640
tgaggcagga ggaacactg agctcagag ttcaaggtt cgtgagcaa tgttcacgc
26700
actgtctcc agcctgattg acaggccaga cctgactct aaacaaaaa aaaaaacaa
26760
tatttaagta atttccaaac atagcagaaa atataagcat ggtttatcac ttgatatga
26820
caccacagc tacttaagat agagtcata attcagtaa ttgttgttg gaaagctaag
26880
gtgcaaccc aagccgctc ttcttaggtg ctctcactg gtgtcatcag ctacagcag
26940
cagagcattg ccaggagcta gctcttccct tcaagaaca aagctttgtt taagagcaca
27000
gtagccaca acttgctctt tctctgag tctcttttat ttccctcctt tcttagggat
27060
cacctgtgtt cttagtattt ggggtcttca ccacaacct gctgtctgga aaacccaaa
27120
ggtagcttc tctctgtac ataaatactt ccaagaacta atgtgtgca agtcaacttt
27180
tggtagctaa gcacagaagt ggctatata ttaagggaan tgacacaaat taacaaaaa
27240
taaacatana agccaaaga aatgtaaaac tattctatgt tcttgaaaca ctcttgact
27300
gtatcagtg tttctttcat gtaagccact aaggtttaag atctattact tgtaacagga
27360
agctggagta tatgtctctg taataattg ccacatcctc attttgactt gatttctaag
27420
tggtgcaca tccatttcta agtggatgta tctccatagt gaataataa ccacttgcaa
27480

WO 02/34922

PCT/US01/42528

tagtatitit gtttgcctgg gtatcagaca aatcagctgt gaagctgcaa ggtctgcagg
27540
tctgaaggta cactgccag tctagtagcc acgggccaca tccgctact gacacatga
27600
catgtggcca gttggaattg agttgtgctg taagttttaa atacgtgctg gattttgaag
27660
acatagtacc ctaaaaaat gtgaacatt tccctttagt aattatttat attgattaca
27720
ggttgggctg gtaatttttg gtttaataaa ctctattaag attaacttca ccttttaaaa
27780
atgtgaccac cagacattt taatttccac atgtagatca ccttatatt ctattgatcg
27840
gtgctagggtg gtgggtgaag aattgtgttc atgtgttttg ggggtggtg ttgggtgtgt
27900
cctctcattt caggcttttg accccttgag gttctctcag gagaattctg atcagagaca
27960
cccttatgcc tacttaccat tctcagctgg atcaaggtag gaacatttg aattgtctga
28020
aagtaccaca agatgtttac ttgagagtag ttatttcctt tccgtctctc agctctatac
28080
attcttccag ggaacgttag ctcttggtgc ctatttgagc ccaaaaggat cagttagttt
28140
tacaaggac aatcgtattc tctgtcacat cctttttggc catgcctcaa aagcagtcac
28200
acattgtatg ctactgtcca taggtctcat gcagtcacc ttcaagcaa gagaataat
28260
ttcatgagta actccactg cgccttggtt atagggaagg catcctgttg gacccctcca
28320
gtcacaattc tccagtgaa caatttaagt cttaagttaa aaglltcaa tggcattttg
28380
tggaaaaaat atcactttac tgtgtacttc agacttcttg tactagtatt ttactatagt
28440
cagaagaanc atcattttt caaglaicac ttctttccc tcttgcttc aggaactgca
28500
ttgggcagga gtttgcctg attgagttta aggttaacct tgccttgatt ctgctccact
28560
tcagagtgac tccagacccc accaggcctc ttactttccc caaccatttt atcctcaagc
28620
ccaagaatgg gatgtatttg caccgaaga aactctctga atgttagatc tcagggtaca
28680
atgattaaac gtactttgtt ttctgaagtt aaatttaccg ctaattgatcc aagcagatag
28740
aagggtatca atgtatggtg ggaggatttg aggttggttg gatagggttc tctgtgaaga
28800
gatccaaat catttctagg tacacagtgt gtcagctaga tctgtttcta tataactttg
28860
ggagattttc agatcttttc tgttaacctt tccactctat taatgctga tacaccaata
28920
gactttcata ttttttctgt tgttttttaa atagttttca gaattatgca agtaataagt
28980
gcatgtatgo tccactgtca aaatttccaa cactagaaaa tcatgtagaa taaaaatttt
29040
aaatctcact tccatttagc gacattccat gccctgacca atcctactgc ttttctaaa
29100
aacagaataa ttgtgtgtgc attttttcag actttttcct atacatttta tatgtagaa
29160
tgtagcaatg tttttgtata gatgtatca ttcctatat ttttttatt tttttcactt
29220
aataaaaatt cactttatc ctatcattg ctttatggtt ttctgtata tgaatgtact
29280

WO 02/34922

PCT/US01/42528

ataattttatt toactatttt cttatttggg catttaagtt atttctagtt ttaaaaacat
29340
gcttgtcaat ggcaacaaaa gccaaaattg acaantggga tctaattaaa ctaaaagagct
29400
tctgcacagc aaacacaaact accatcacac tgaatgggca gccacagaa tgggagaaaa
29460
tttttgcaac ctactcatct gacaaaggcc taatatccag attctacaat gaactcaaac
29520
aaatgtcacg gaaaaaaca acccatcaa aaagtgggtg aaggatatga acagacactt
29580
ctcaaaagaa gacattttacg cagccaaaag acacatgaaa aaatgcctat cgtcactggc
29640
catcagagaa atgcaaatca aaaccacaat gagataccat ctccaccag ttagaatggc
29700
aatcattaaa aagtcaggaa acaacagggtg ctggagagga tgtggagaaa taggaagact
29760
tttacaactgt tgggtggcagg agaactcatt gaacccggga gggggaggtt gcagtgaacc
29820
gaggtggcgc cactgcactc cagcctgggc gacagaacga gtactccatc tcaaaaaaaa
29880
aaaaaaaggc caccacactt ctcaatctta atgttgcct ctatgtggtt tcttccata
29940
tctctctcag acagagtcct cttttgctga tatgatctta cagtattttt tgtttatacc
30000
attataatct cattaatgac agcaacaaa atgacaaaag caactgatt tctccccttg
30060
gatgacctaa ttgtcttca ctcttccatc atcacttata acatgatgat tctcaattc
30120
ctctccttaa atctctata taaaaaatic cctcccttga attccagatc ctgggagaca
30180
aacacccacg tctaaaccaa aatttggtta acactggacc agtcgtcctg tgtgacttcc
30240
cattttgtgc ctattttgtc agctgggtata ccaatatcca cccagttana caatatttcc
30300
ttgttttttt ctggtacaaa cccaataaa ttacaaact caataaaagt aaatttctaa
30360
ataaactcac ttctctata tatctccttc ttgttggaan aatgggttag gtagttctt
30420
taaaagcatg catgataaat tgaatgaat acaatttca ggtcggaca tactaggat
30480
aattttctgt gtctctgggg tcttacctat ttggggtcac aataaacaag tttattttag
30540
ttattaatat tcaatttcat tatcttcttt acaattatg ttccctggtt gtttcatlga
30600
caataattta ttgtcaggt tgcaggtgc ttctaacctt ctgtgtattt ttcatatcc
30660
aattttactt taatatattt tagaaaagag gtctgtttaa ttctctaata attattatat
30720
tatgtttttt tcaatgacat ttgtgaatt gaaaacccctt aaaaatatga aatcattttt
30780
tcgaactctg tgcacacagc aattttgtta aataagaaga cagaacagg gcattatcaa
30840
gagataaata ttcaatatac cttatatttc tgtcacacat ttttatacca aotgtgcaa
30900
aaattgtata tcatataaat gatacaagt tcacaaaggc attcctttat ccttaactc
30960
tcaaataga aactttcata ggtagggaagt aggggaagca tatattccct ttgaagggtg
31020
caagaaaatg tcaattggcat tcaccatggt actcttcaag cttaaaaaaa atggactgca
31080

WO 02/34922

PCT/US01/42528

aaacatttac aaacatagca tatttattgg gtacotttat gttacataa atattgaaga
31140
tatctcacat acctctttca atcagattat ctccctgaca ttatttgacc actttctatg
31200
gggaasac
31208

<210> 4
<211> 489
<212> PRT
<213> Human

<400> 4
Val Ala Ala Leu Leu Gly Leu Leu Leu Leu Leu Lys Ala Ala Gln
1 5 10 15
Leu Tyr Leu His Arg Gln Trp Leu Leu Arg Ala Leu Gln Gln Phe Pro
20 25 30
Cys Pro Pro Phe His Trp Leu Leu Gly His Ser Arg Glu Phe Gln Asn
35 40 45
Asp Gln Glu Leu Glu Arg Ile Gln Lys Trp Val Glu Lys Phe Pro Gly
50 55 60
Ala Cys Pro Trp Trp Leu Ser Gly Asn Lys Ala Arg Leu Leu Val Tyr
65 70 75 80
Asp Pro Asp Tyr Leu Lys Val Ile Leu Gly Arg Ser Asp Pro Lys Ala
85 90 95
Pro Arg Asn Tyr Lys Leu Met Thr Pro Trp Ile Gly Tyr Gly Leu Leu
100 105 110
Leu Leu Asp Gly Gln Thr Trp Phe Gln His Arg Arg Met Leu Thr Pro
115 120 125
Ala Phe His Tyr Asp Ile Leu Lys Pro Tyr Val Gly Leu Met Val Asp
130 135 140
Ser Val Gln Ile Met Leu Asp Arg Trp Glu Gln Leu Ile Ser Gln Asp
145 150 155 160
Ser Ser Leu Glu Ile Phe Gln His Val Ser Leu Met Thr Leu Asp Thr
165 170 175
Ile Met Lys Cys Ala Phe Ser Tyr Gln Gly Ser Val Gln Leu Asp Arg
180 185 190
Asn Ser His Ser Tyr Ile Gln Ala Ile Asn Asp Leu Asn Asn Leu Val
195 200 205
Phe Tyr Arg Ala Arg Asn Val Phe His Gln Ser Asp Phe Leu Tyr Arg
210 215 220
Leu Ser Pro Glu Gly Arg Leu Phe His Arg Ala Cys Gln Leu Ala His
225 230 235 240
Glu His Thr Asp Arg Val Ile Gln Gln Arg Lys Ala Gln Leu Gln Gln
245 250 255
Glu Gly Glu Leu Glu Lys Val Arg Arg Lys Arg Arg Leu Asp Phe Leu
260 265 270
Asp Val Leu Leu Phe Ala Lys Met Glu Asn Gly Ser Ser Leu Ser Asp
275 280 285
Gln Asp Leu Arg Ala Glu Val Asp Thr Phe Met Phe Glu Gly His Asp
290 295 300
Thr Thr Ala Ser Gly Val Ser Trp Ile Phe Tyr Ala Leu Ala Thr His
305 310 315 320
Pro Glu His Gln His Arg Cys Arg Glu Glu Ile Gln Gly Leu Leu Gly
325 330 335
Asp Gly Ala Ser Ile Thr Trp Glu His Leu Asp Gln Met Pro Tyr Thr
340 345 350
Thr Met Cys Ile Lys Glu Ala Leu Arg Leu Tyr Pro Pro Val Pro Ser
355 360 365

WO 02/34922

PCT/US01/42528

```

Val Thr Arg Gln Leu Ser Lys Pro Val Thr Phe Pro Asp Gly Arg Ser
370          375          380
Leu Pro Lys Gly Val Ile Leu Phe Leu Ser Ile Tyr Gly Leu His Tyr
385          390          395          400
Asn Pro Lys Val Trp Gln Asn Pro Glu Val Phe Asp Pro Phe Arg Phe
405          410          415
Ala Pro Asp Ser Ala Tyr His Ser His Ala Phe Leu Pro Phe Ser Gly
420          425          430
Gly Ala Arg Asn Cys Ile Gly Lys Gln Phe Ala Met Arg Glu Leu Lys
435          440          445
Val Ala Val Ala Leu Thr Leu Leu Arg Phe Glu Leu Leu Pro Asp Pro
450          455          460
Thr Arg Val Pro Ile Pro Ile Ala Arg Val Val Leu Lys Ser Lys Asn
465          470          475          480
Gly Ile His Leu Arg Leu Arg Lys Leu
485

```


【国際公開パンフレット（コレクトバージョン）】

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
2 May 2002 (02.05.2002)

PCT

(10) International Publication Number
WO 02/034922 A3

- (51) International Patent Classification: C12N 15/53, 9/02, C07K 16/40, C12Q 1/68, G01N 33/68
- (21) International Application Number: PCT/US01/42528
- (22) International Filing Date: 5 October 2001 (05.10.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/241,745 20 October 2000 (20.10.2000) US
09/739,456 19 December 2000 (19.12.2000) US
09/818,647 28 March 2001 (28.03.2001) US
09/852,067 10 May 2001 (10.05.2001) US
- (71) Applicant: PE CORPORATION (NY) (US/US); 761 Main Avenue, Norwalk, CT 06859 (US).
- (72) Inventors: MERKULOV, Gennady, V.; c/o Celera Genomics, 45 West Gude Drive C2-4#20, Rockville, MD 20850 (US); YAN, Chunhui; c/o Celera Genomics, 45 West Gude Drive C2-4#20, Rockville, MD 20850 (US); DI FRANCESCO, Valentina; c/o Celera Genomics, 45 West Gude Drive C2-4#20, Rockville, MD 20850 (US); BEASLEY, Ellen, M.; c/o Celera Genomics, 45 West Gude Drive C2-4#20, Rockville, MD 20850 (US).
- (74) Agent: MILLMAN, Robert, A.; Celera Genomics, Chief Intellectual Property Counsel, 45 West Gude Drive C2-4#20, Rockville, MD 20850 (US).
- (81) Designated States (national): AU, AC, A1, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GR, GU, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MY, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GI, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:
with international search report
- (85) Date of publication of the international search report: 28 August 2003
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 02/034922 A3

(54) Title: ISOLATED HUMAN DRUG-METABOLIZING PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN DRUG-METABOLIZING PROTEINS, AND USES THEREOF

(57) Abstract: The present invention provides amino acid sequences of peptides that are encoded by genes within the human genome, the drug-metabolizing enzyme peptides of the present invention. The present invention specifically provides isolated peptide and nucleic acid molecules, methods of identifying orthologs and paralogs of the drug-metabolizing enzyme peptides, and methods of identifying modulators of the drug-metabolizing enzyme peptides.

【国際調査報告】

| | | |
|---|---|--|
| INTERNATIONAL SEARCH REPORT | | International application No. PCT/US 01/42528 |
| A. CLASSIFICATION OF SUBJECT MATTER IPC 7 C12N15/53 C12N9/02 C07K16/40 C12Q1/68 G01N33/68 | | |
| According to International Patent Classification (IPC) or to both national classification and IPC | | |
| B. REIDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC 7 C12N C07K C12Q | | |
| Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched | | |
| Electronic data base consulted during the international search (name of data base and, where practical, search terms used) EPO-Internal, SEQUENCE SEARCH, BIOSIS, EMBL, WPI Data, PAJ | | |
| C. DOCUMENTS CONSIDERED TO BE RELEVANT | | |
| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
| A | MATSUBARA S ET AL: "COMPLEMENTARY DNA CLONING AND INDUCIBLE EXPRESSION DURING PREGNANCY OF THE MESSENGER RNA FOR RABBIT PULMONARY PROSTAGLANDIN OMEGA HYDROXYLASE CYTOCHROME P-450-P-2" JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 262, no. 27, 1987, pages 13366-13371, XP002241281 ISSN: 0021-9258 the whole document | 1-23 |
| A | WO 95 30766 A (US HEALTH) 16 November 1995 (1995-11-16) -/- | |
| <input checked="" type="checkbox"/> Further documents are listed in the continuation of box C. <input checked="" type="checkbox"/> Patent family members are listed in annex. | | |
| * Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "S" document member of the same patent family | | |
| Date of the actual completion of the international search 15 May 2003 | | Date of mailing of the international search report 23 05 2003 |
| Name and mailing address of the ISA European Patent Office, P.B. 5618 Patensan 2 NL - 2280 HV Rijswijk Tel: (+31-70) 340-2040, Tx: 31 851 epo nl, Fax: (+31-70) 340-9010 | | Authorized officer Le Cornec, N |

INTERNATIONAL SEARCH REPORT

 International Application No.
 PCT/US 01/42528

| C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT | | |
|---|---|-----------------------|
| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
| A | DATABASE EMBL [Online] 15 March 1998 (1998-03-15) NCI-CGAP: "oh04f03.s1 NCI CGAP KID3 Homo sapiens cDNA clone IMAGE:1456829 3' similar to SW:CP48 RAT P24464 CYTOCHROME P450 IVA8; mRNA sequence." retrieved from EBI, HINXTON, UK Database accession no. AA863360 XP002241282 abstract & UNPUBLISHED. | 1-5 |
| A | --- DATABASE EMBL [Online] 11 September 1997 (1997-09-11) NCI-CGAP: "n181a02.s1 NCI CGAP-Br2 Homo sapiens cDNA clone IMAGE:1057034 3' mRNA sequence" retrieved from EBI, HINXTON, UK Database accession no. AA557324 XP002241283 abstract & UNPUBLISHED. | 1-5 |
| P,X | --- WO 01 51638 A (INCYTE GENOMICS INC ;AZIMZAI YALDA (US); REDDY ROOPA (US); RING HU) 19 July 2001 (2001-07-19) The whole document especially sequences ID no.6 and 38 | 1-23 |
| P,X | --- WO 01 40466 A (STEWART TIMOTHY A ;BAKER KEVIN P (US); DEFORGE LAURA (US); DESNOYE) 7 June 2001 (2001-06-07) sequences ID no.107 and 108 claims | 1-23 |
| L | --- DATABASE GENESEQ [Online]--- 17 December 2001 (2001-12-17) retrieved from EBI, HINXTON, UK Database accession no. AAS40801 XP002241284 L document cited to provide information on the relevant sequence ID no.27 disclosed in WO 01 55301 the whole document | 1-5 |
| P,X | --- -& WO 01 55301 A (HUMAN GENOME SCIENCES INC ;ROSEN CRAIG A (US); BARASH STEVEN C (US) 2 August 2001 (2001-08-02) | |
| E | --- WO 01 77137 A (HASELTINE WILLIAM A ;HUMAN GENOME SCIENCES INC (US); ROSEN CRAIG A) 18 October 2001 (2001-10-18) sequence ID no.1760 claims | 1-23 |
| | --- -/-- | |

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

| INTERNATIONAL SEARCH REPORT | | International Application No. PCT/US 01/42528 |
|---|--|--|
| C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT | | |
| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
| E | WO 02 16388 A (HUMAN GENOME SCIENCES INC ;BAKER KEVIN P (US); NI JIAN (US); ROSEN) 28 February 2002 (2002-02-28) sequences ID no.16 and 66 claims page 20 -page 25 ----- | 1-23 |

INTERNATIONAL SEARCH REPORT

International Application No.
PCT/JP 01/42528

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
Although claim 18 is directed to a method of treatment of the human/animal body (rule 39.1 (IV) PCT), the search has been carried out and based on the alleged effects of the compound/composition.
2. ☒ Claims Nos.: 17, 18 partially
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this International application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

International Application No. PCT/US 01/42528

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 17, 18 partially

Claim 17 refers to a pharmaceutical composition comprising an agent identified by the method of claim 16 and claim 18 refers to a method for treating a disease by administering an agent identified by the method of claim 16 without giving a true technical characterization. Moreover, no such compounds are defined in the application. In consequence, the scope of said claims is ambiguous and vague, and their subject-matter is not sufficiently disclosed and supported. No search can be carried out for such purely speculative claims whose wording is, in fact, a mere recitation of the results to be achieved. But a partial search has been carried out as far as the agent is an antibody against the polypeptide, a ribozyme a probe or an antisense.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US 01/42528

| Patent document cited in search report | | Publication date | | Patent family member(s) | Publication date |
|---|---|---------------------|----|----------------------------|---------------------|
| WO 9530766 | A | 16-11-1995 | US | 5912120 A | 15-06-1999 |
| | | | AU | 2476695 A | 29-11-1995 |
| | | | WO | 9530766 A1 | 16-11-1995 |
| WO 0151638 | A | 19-07-2001 | AU | 2790001 A | 24-07-2001 |
| | | | CA | 2397340 A1 | 19-07-2001 |
| | | | EP | 1254235 A2 | 06-11-2002 |
| | | | WO | 0151638 A2 | 19-07-2001 |
| WO 0140466 | A | 07-06-2001 | AU | 2055401 A | 12-06-2001 |
| | | | AU | 2192800 A | 12-07-2000 |
| | | | AU | 2474700 A | 19-06-2000 |
| | | | AU | 2879400 A | 31-07-2001 |
| | | | AU | 3514400 A | 28-09-2000 |
| | | | AU | 3774300 A | 18-12-2000 |
| | | | AU | 5152700 A | 02-01-2001 |
| | | | AU | 6391000 A | 19-02-2001 |
| | | | CA | 2347677 A1 | 08-06-2000 |
| | | | CA | 2353775 A1 | 29-06-2000 |
| | | | CA | 2362427 A1 | 14-09-2000 |
| | | | CA | 2365610 A1 | 26-07-2001 |
| | | | CA | 2372511 A1 | 21-12-2000 |
| | | | CA | 2383254 A1 | 07-12-2000 |
| | | | CA | 2391455 A1 | 07-06-2001 |
| | | | EP | 1173563 A1 | 23-01-2002 |
| | | | EP | 1220905 A2 | 10-07-2002 |
| | | | EP | 1210418 A1 | 05-06-2002 |
| | | | EP | 1208195 A2 | 29-05-2002 |
| | | | EP | 1250426 A2 | 23-10-2002 |
| | | | EP | 1141289 A2 | 10-10-2001 |
| | | | EP | 1135495 A2 | 26-09-2001 |
| | | | US | 2003032057 A1 | 13-02-2003 |
| | | | WO | 0153486 A1 | 26-07-2001 |
| | | | WO | 0053758 A2 | 14-09-2000 |
| | | | WO | 0073454 A1 | 07-12-2000 |
| | | | WO | 0077037 A2 | 21-12-2000 |
| | | | WO | 0109327 A2 | 08-02-2001 |
| | | | WO | 0140466 A2 | 07-06-2001 |
| | | | US | 2003004311 A1 | 02-01-2003 |
| | | | US | 2002127584 A1 | 12-09-2002 |
| | | | US | 2003044902 A1 | 06-03-2003 |
| | | | US | 2003044844 A1 | 06-03-2003 |
| | | | US | 2003040014 A1 | 27-02-2003 |
| | | | US | 2003032062 A1 | 13-02-2003 |
| | | | US | 2003032063 A1 | 13-02-2003 |
| | | | US | 2002177165 A1 | 28-11-2002 |
| | | | WO | 0037640 A2 | 29-06-2000 |
| | | | US | 2003073210 A1 | 17-04-2003 |
| | | | US | 2003022239 A1 | 30-01-2003 |
| | | | US | 2003054516 A1 | 20-03-2003 |
| | | | US | 2003073211 A1 | 17-04-2003 |
| | | | US | 2003073212 A1 | 17-04-2003 |
| | | | US | 2003022328 A1 | 30-01-2003 |
| | | | US | 2003073213 A1 | 17-04-2003 |
| | | | US | 2003073214 A1 | 17-04-2003 |
| | | | US | 2003032155 A1 | 13-02-2003 |
| | | | US | 2003022331 A1 | 30-01-2003 |

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No.
PCT/US 2001/02528

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|--|------------------|-------------------------|-------------------------|
| WO 0155301 | A | 02-08-2001 | AU 2950801 A 07-08-2001 |
| | | | AU 3095801 A 07-08-2001 |
| | | | AU 3645901 A 07-08-2001 |
| | | | AU 3646001 A 07-08-2001 |
| | | | AU 3646101 A 07-08-2001 |
| | | | AU 3646201 A 07-08-2001 |
| | | | AU 3646301 A 07-08-2001 |
| | | | AU 3646401 A 07-08-2001 |
| | | | AU 3646501 A 07-08-2001 |
| | | | AU 3646601 A 07-08-2001 |
| | | | AU 3794301 A 07-08-2001 |
| | | | AU 3794401 A 07-08-2001 |
| | | | AU 3794701 A 07-08-2001 |
| | | | AU 3794901 A 07-08-2001 |
| | | | AU 3795001 A 07-08-2001 |
| | | | AU 3795101 A 07-08-2001 |
| | | | AU 3795201 A 07-08-2001 |
| | | | AU 3795301 A 07-08-2001 |
| | | | AU 3795401 A 07-08-2001 |
| | | | AU 3795501 A 07-08-2001 |
| | | | AU 3795701 A 07-08-2001 |
| | | | AU 3795801 A 07-08-2001 |
| | | | AU 3972601 A 07-08-2001 |
| | | | AU 3972701 A 07-08-2001 |
| | | | AU 3972801 A 07-08-2001 |
| | | | AU 4140201 A 07-08-2001 |
| | | | AU 4140301 A 07-08-2001 |
| | | | AU 4140401 A 07-08-2001 |
| | | | AU 4140501 A 07-08-2001 |
| | | | AU 4140601 A 07-08-2001 |
| | | | AU 4140701 A 07-08-2001 |
| | | | AU 4140801 A 07-08-2001 |
| | | | AU 4140901 A 07-08-2001 |
| | | | AU 4141001 A 07-08-2001 |
| | | | AU 4141101 A 20-08-2001 |
| | | | AU 4141201 A 07-08-2001 |
| | | | AU 4141301 A 07-08-2001 |
| | | | AU 4141401 A 07-08-2001 |
| | | | AU 4141501 A 07-08-2001 |
| | | | AU 4141601 A 07-08-2001 |
| | | | AU 4141701 A 07-08-2001 |
| | | | AU 4141801 A 07-08-2001 |
| | | | AU 4141901 A 07-08-2001 |
| | | | AU 4313401 A 07-08-2001 |
| | | | AU 4313501 A 07-08-2001 |
| | | | AU 4313601 A 07-08-2001 |
| | | | AU 4313701 A 14-08-2001 |
| | | | AU 4526201 A 07-08-2001 |
| | | | AU 4719001 A 07-08-2001 |
| | | | AU 4719101 A 07-08-2001 |
| WO 0177137 | A | 18-10-2001 | AU 5906301 A 30-10-2001 |
| | | | AU 5906601 A 30-10-2001 |
| | | | AU 6102401 A 30-10-2001 |
| | | | AU 6294201 A 30-10-2001 |
| | | | AU 6456301 A 30-10-2001 |
| | | | AU 6655701 A 23-10-2001 |
| | | | AU 7480901 A 30-10-2001 |

INTERNATIONAL SEARCH REPORT

Indicate on patent family members

International class No
PCT/US 01/42528

| Patent document cited in search report | | Publication date | Patent family member(s) | Publication date |
|---|---|---------------------|----------------------------|---------------------|
| WO 0177137 | A | | CA 2405525 A1 | 25-10-2001 |
| | | | CA 2405550 A1 | 25-10-2001 |
| | | | CA 2405563 A1 | 25-10-2001 |
| | | | CA 2405701 A1 | 25-10-2001 |
| | | | CA 2405709 A1 | 25-10-2001 |
| | | | EP 1274719 A2 | 15-01-2003 |
| | | | EP 1274720 A1 | 15-01-2003 |
| | | | EP 1278767 A1 | 29-01-2003 |
| | | | EP 1276856 A1 | 22-01-2003 |
| | | | EP 1276849 A2 | 22-01-2003 |
| | | | EP 1278544 A2 | 29-01-2003 |
| | | | EP 1276756 A1 | 22-01-2003 |
| | | | WO 0179442 A2 | 25-10-2001 |
| | | | WO 0179443 A2 | 25-10-2001 |
| | | | WO 0177137 A1 | 18-10-2001 |
| | | | WO 0179480 A1 | 25-10-2001 |
| | | | WO 0179258 A1 | 25-10-2001 |
| | | | WO 0179271 A1 | 25-10-2001 |
| | | | WO 0179444 A2 | 25-10-2001 |
| WO 0216388 | A | 28-02-2002 | AU 2950701 A | 04-03-2002 |
| | | | WO 0216388 A1 | 28-02-2002 |

フロントページの続き

| (51)Int.Cl. ⁷ | F I | テーマコード (参考) |
|--------------------------|----------------|-------------|
| C 1 2 M 1/00 | C 1 2 M 1/00 | A 4 C 0 8 4 |
| C 1 2 N 1/15 | C 1 2 N 1/15 | 4 H 0 4 5 |
| C 1 2 N 1/19 | C 1 2 N 1/19 | |
| C 1 2 N 1/21 | C 1 2 N 1/21 | |
| C 1 2 N 5/10 | C 1 2 N 9/00 | |
| C 1 2 N 9/00 | C 1 2 Q 1/02 | |
| C 1 2 Q 1/02 | C 1 2 Q 1/25 | |
| C 1 2 Q 1/25 | C 1 2 Q 1/68 | A |
| C 1 2 Q 1/68 | G O 1 N 33/53 | D |
| G O 1 N 33/53 | G O 1 N 33/53 | M |
| G O 1 N 33/566 | G O 1 N 33/566 | |
| G O 1 N 37/00 | G O 1 N 37/00 | 1 O 2 |
| | C 1 2 N 15/00 | F |
| | C 1 2 N 5/00 | A |

(31)優先権主張番号 09/852,067

(32)優先日 平成13年5月10日(2001.5.10)

(33)優先権主張国 米国(US)

(81)指定国 AP(GH,GM,KE,LS,MW,MZ,SD,SL,SZ,TZ,UG,ZW),EA(AM,AZ,BY,KG,KZ,MD,RU,TJ,TM),EP(AT,BE,CH,CY,DE,DK,ES,FI,FR,GB,GR,IE,IT,LU,MC,NL,PT,SE,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AT,AU,AZ,BA,BB,BG,BR,BY,BZ,CA,CH,CN,CO,CR,CU,CZ,DE,DK,DM,DZ,EC,EE,ES,FI,GB,GD,GE,GH,GM,HR,HU,ID,IL,IN,IS,JP,KE,KG,KP,KR,KZ,LC,LK,LR,LS,LT,LU,LV,MA,MD,MC,MK,MN,MW,MX,MZ,NO,NZ,PH,PL,PT,R O,RU,SD,SE,SG,SI,SK,SL,TJ,TM,TR,TT,TZ,UA,UG,UZ,VN,YU,ZA,ZW

(特許庁注：以下のものは登録商標)

Windows

(72)発明者 ヤン, クーンホア

アメリカ合衆国 メリーランド州 20850 ロックビル, C2-4#21, ウェスト グーデ
ドライブ 45, セレーラ ジェニミクス内

(72)発明者 ディ フランセスコ, バレンティナ

アメリカ合衆国 メリーランド州 20850 ロックビル, C2-4#21, ウェスト グーデ
ドライブ 45, セレーラ ジェニミクス内

(72)発明者 ビーズリー, エレン, エム

アメリカ合衆国 メリーランド州 20850 ロックビル, C2-4#21, ウェスト グーデ
ドライブ 45, セレーラ ジェニミクス内

Fターム(参考) 4B024 AA01 AA11 AA19 BA07 BA80 CA03 CA04 CA09 CA12 CA20
DA01 DA02 DA05 DA11 DA12 GA11 HA11 HA13 HA14
4B029 AA07 AA21 AA23 BB20 CC01 CC02 CC03 CC08 FA12 FA15
4B050 CC01 CC03 DD11 EE01 LL01 LL03 LL05
4B063 QA01 QA05 QQ02 QQ08 QQ21 QQ41 QQ43 QQ53 QQ61 QQ89
QQ95 QR01 QR08 QR32 QR35 QR40 QR42 QR48 QR56 QR62
QR77 QR80 QR84 QS16 QS25 QS33 QS34 QS36 QX01 QX02
4B065 AA01X AA58X AA72X AA87X AA93Y AB01 AC14 BA02 CA27 CA43
CA44 CA46
4C084 AA17 NA05 NA06 NA14 ZC202 ZC412

(193)

JP 2004-531207 A 2004.10.14

4H045 AA11 CA40 DA75 EA20 EA50 FA71